BALKAN ANALYTIC FORUM

BAF2: Dispositions
BAF+: Dispositions and Values

Edited by
Miroslava Trajkovski
University of Belgrade
Emily C. McWilliams
University of Dayton

1838
UNIVERSITY OF BELGRADE
FACULTY OF PHILOSOPHY

Welcome to the second volume of the Balkan Analytic Forum's proceedings, featuring contributions from *BAF2: Dispositions* and *BAF2+: Dispositions and Values*. It is a pleasure and honor to co-edit these proceedings with Dr. Miroslava Trajkovski for the second consecutive year. This volume showcases the vibrant intellectual community and the spirit of curiosity that animate the *Balkan Analytic Forum*. Now entering its third year, the Forum continues to foster regional philosophical community and international scholarly exchange while engaging global philosophical conversations. Rooted in analytic philosophy, it also seeks to bridge traditions, welcoming approaches that connect analytic methods with phenomenological, historical, and other philosophical perspectives and traditions.

Reflecting this commitment to intellectual breadth, the essays collected here engage the concept of dispositions through a variety of philosophical methods and traditions. They explore dispositions as epistemological, ontological, and ethical phenomena, theorizing their nature and function within domains including evidentiary reasoning, cosmological theorizing, virtue ethics, aesthetic experience, and philosophy of mind.

The Forum's vibrant intellectual community owes immense gratitude to Dr. Miroslava Trajkovski, whose leadership and unwavering commitment have created a collaborative space where philosophical curiosity thrives. The Balkans' rich intellectual history reminds us that philosophical vitality often emerges at the intersection of traditions. The Forum carries that spirit forward in the work collected here.

Emily C. McWilliams

University of Belgrade – Faculty of Philosophy | 2025

1838

Balkan Analytic Forum
https://www.f.bg.ac.rs/balkanski_analiticki_forum
Dispositions & Dispositions and Values
Conference Proceedings
10–20. X 2024. Belgrade, Serbia

Edited by
MIROSLAVA TRAJKOVSKI, University of Belgrade (0000-0002-4927-2168)
EMILY C. McWILLIAMS, University of Dayton (0000-0002-8609-070X)

Publisher
Center for Contemporary Philosophy – Balkan Analytic Forum
University of Belgrade – Faculty of Philosophy
Čika Ljubina 18–20, Beograd 11000, Srbija
www.f.bg.ac.rs

For the publisher
Prof. Dr. Danijel Sinani
Dean of the Faculty of Philosophy

Cover art and design by
Dragan Trajkovski

Set by
Dosije studio, Belgrade

Supported by:
MINISTRY OF SCIENCE, TECHNOLOGICAL
DEVELOPMENT AND INNOVATION
of the Republic of Serbia

Република Србија
МИНИСТАРСТВО НАУКЕ,
ТЕХНОЛОШКОГ РАЗВОЈА И ИНОВАЦИЈА

Printed by
JP Službeni glasnik

Print run
200

Project number 451-03-137/2025-03/200163

ISBN 978-86-6427-381-7

https://doi.org/10.18485/baf.2025.2

Balkan Analytic Forum

Dispositions
&
Dispositions and Values

Conference Proceedings
10–20. X 2024. Belgrade, Serbia

Edited by

MIROSLAVA TRAJKOVSKI
University of Belgrade

EMILY C. McWILLIAMS
University of Dayton

Belgrade
2025

CONTENTS

7 | Contributors

9 | Preface
Emily C. McWilliam

10 | Acknowledgements

11 | Hybridity as Authenticity: Cultural Contact, Geography and Early Modern Balkan Visual Culture – *Nenad Makuljević*

17 | Introduction: Dispositions & Dispositions and Values – *Miroslava Trajkovski*

BAF2: DISPOSITIONS

1. DISPOSITIONS AND EVIDENCE

33 | *Timothy Williamson*, “Recognitional Capacities and their Uses”

49 | *Slobodan Perović*, “The Copernican Principle and Evidence in Modern Cosmology”

2. DISPOSITIONS AND PERSONALITY TRAITS

63 | *Aleksandra M. Pavlović*, “Dispositions and Brain Health: The Role of Premorbid Personality Traits in Stroke Risk Assessment”

81 | *Goran Knežević*, “Major Dispositions in the Psychology of Individual Differences: Conceptualization, Measurement, Origins, and Consequences”

3. EPISTEMIC DISPOSITIONS

99 | *Iris Vidmar Jovanović*, “Epistemic Dispositions of Literature: Insights from Literary Tradition and Critical Practice”

121 | *Ana Kuburić Zotova*, “How mindfulness as a disposition impacts epistemic dispositions”

145 | *Ognjen Milivojević*, "Background and Abilities as Dispositions"

4. DISPOSITIONS AND ONTOLOGY

157 | *Una Popović and Srđan Šarović*, "Donkey's dilemma: values or valor?"

169 | *Dušan Smiljanić*, "Dispositions – A Property or a Way of Being"

BAF2+: DISPOSITIONS AND VALUES

5. DISPOSITIONS AND VIRTUES

193 | *Amber Riaz*, "Moral Learning for the Wretched of the Earth"

209 | *Marcin Trepczyński*, "Virtues as Dispositions: Different Approaches in Medieval Analytic Thought"

6. DISPOSITIONS AND EMOTIONS

231 | *Damir Smiljanić*, "Cool philosophy or the art of restraining from judging without being indifferent"

245 | *Isidora Novaković*, "Dispositions to tragic emotions"

7. DISPOSITIONS IN AESTHETICS AND ETHICS

267 | *Aleksandar Risteski*, "The Anthropological Implications of Peirce's Aesthetics"

285 | *Natalia Tomashpolskaia,* "Ludwig Wittgenstein's critique of the dispositional theory of values"

CONTRIBUTORS

- Emily C. McWilliams, University of Dayton, Department of Philosophy, Fitz Center for Leadership in Community, USA (0000-0002-8609-070X)
- Miroslava Trajkovski, University of Belgrade, Department of Philosophy, Balkan Analytic Forum (BAF), Serbia (0000-0002-4927-2168)
- Nenad Makuljević, University of Belgrade, Department of Art History, Center for Visual Culture of the Balkans, Serbia (0000-0003-1875-2799)
- Timothy Williamson, New College, University of Oxford, United Kingdom (0000-0002-4659-8672)
- Slobodan Perović, University of Belgrade, Department of Philosophy, Serbia (0000-0003-4393-4105)
- Aleksandra M. Pavlović University of Belgrade, Faculty of Special Education and Rehabilitation, Serbia (0000-0002-5987-9828)
- Goran Knežević, University of Belgrade, Department of Psychology and Laboratory for the research of individual differences (LIRA), Serbia (0000-0001-8951-3774)
- Iris Vidmar Jovanović, Department of Philosophy, Faculty of Humanities and Social Sciences, University of Rijeka, Croatia (0000-0003-3780-6457)
- Ana Kuburić Zotova, independent researcher, University of Belgrade, Serbia (0000-0002-2592-9831)
- Ognjen Milivojević, graduate student, University of Belgrade, Department of Philosophy, Serbia (0009-0007-6696-8144)
- Una Popović, University of Novi Sad, Department of Philosophy, Serbia (0000-0002-3024-1345)
- Srđan Šarović, University of Novi Sad, Academy of Arts, Serbia (0009-0004-1255-3558)

- Dušan Smiljanić, MA student, University of Belgrade, Serbia (0000-0003-0673-338X)
- Amber Riaz, Lahore University of Management Sciences: Lahore, Pakistan (0009-0002-1142-2404)
- Marcin Trepczyński, University of Warsaw, Faculty of Philosophy, Poland (0000-0003-0612-2597)
- Damir Smiljanić, University of Novi Sad, Department of Philosophy, Serbia (0000-0002-5791-620X)
- Isidora Novaković, graduate student, University of Belgrade, Department of Philosophy, Serbia (0009-0007-2129-5580)
- Natalia Tomashpolskaia, independent researcher, University of Málaga, Spain (0000-0002-3017-4058)
- Aleksandar Risteski, University of Priština in Kosovska Mitrovica, Department of Philosophy, Serbia (0000-0001-5179-3977)

PREFACE

Welcome to the second volume of the Balkan Analytic Forum's proceedings, featuring contributions from BAF2: *Dispositions* and BAF2+: *Dispositions and Values*.

It is a pleasure and honor to co-edit these proceedings with Dr. Miroslava Trajkovski for the second consecutive year. This volume showcases the vibrant intellectual community and the spirit of curiosity that animate the Balkan Analytic Forum. Now entering its third year, the Forum continues to foster regional philosophical community and international scholarly exchange while engaging global philosophical conversations. Rooted in analytic philosophy, it also seeks to bridge traditions, welcoming approaches that connect analytic methods with phenomenological, historical, and other philosophical perspectives and traditions.

Reflecting this commitment to intellectual breadth, the essays collected here engage the concept of dispositions through a variety of philosophical methods and traditions. They explore dispositions as epistemological, ontological, and ethical phenomena, theorizing their nature and function within domains including evidentiary reasoning, cosmological theorizing, virtue ethics, aesthetic experience, and philosophy of mind.

The Forum's vibrant intellectual community owes immense gratitude to Dr. Miroslava Trajkovski, whose leadership and unwavering commitment have created a collaborative space where philosophical curiosity thrives. The Balkans' rich intellectual history reminds us that philosophical vitality often emerges at the intersection of traditions. The Forum carries that spirit forward in the work collected here.

Emily C. McWilliams

ACKNOWLEDGEMENTS

Grateful acknowledgments to the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and the University of Belgrade – Faculty of Philosophy who supported this event.

Grateful acknowledgments to Nenad Makuljević, professor of art history for his willingness to give his introductory text about the Balkans, to Djordje Vukašinović for his help in the realization of online presentations. Thanks to Dragan for his design, his gift, to the BAF.

Nenad Makuljević

HYBRIDITY AS AUTHENTICITY: CULTURAL CONTACT, GEOGRAPHY AND EARLY MODERN BALKAN VISUAL CULTURE[1]

One thing for certain cannot be associated with the Balkans: boredom. Nobody can state, for better or worse, that the Balkans are boring. Whether we talk about history, politics, identities–national or regional–, sport, conspiracy theories and – why not – philosophy, the Balkans always stand out from the rest of Europe as still somewhat mystical, almost mythical and incomprehensible jumble of small nation-states, and regions of extremes and opposites.

During the early modern era, the Balkan Peninsula was a meeting place of different cultures and visual practices. The Ottoman conquests of this territory, from the fourteenth to the sixteenth century, have influenced its culture to develop in different circumstances compared to other parts of Europe, and have impacted the construction of a specific cultural model there. With the Ottoman state, Islamization also arrived, so the Balkan region underwent a significant cultural change. The Balkans have become a European territory with a strong and enduring presence of Islam.[2] The previous Christian art and visual culture, nurtured during the Middle Ages most often in the framework of the Byzantine system and the independent states – such as Bulgaria or Serbia, were suppressed by the erection of numerous Islamic religious structures, such as mosques, türbe mausoleums, and madrasas. The change was most visible in urban centres, such as Thessaloniki, Sarajevo, Skopje or Belgrade, in which Islamic public identity dominated, as clearly evidenced by Evliya Çelebi's travelogues (1611–1682).[3]

1 This work in its complete form was published as: N. Makuljević, Hybridity as Authenticity: Cultural Contact, Geography and Early Modern Balkan Visual Culture, in: *Proceedings of the 34th World Congress of Art History*, vol. 3, edited by: D. Shao– D.Fan– Q. Zhu, Beijing: The Commercial Press 2019, 1514–1518.

2 Leften S. Stavrianos, *The Balkans since 1453*, (Rinehart & Company 1958), 81–105.

3 Evlija Čelebi, Putopis, *Odlomci o jugoslavenskim zemljama*, prevod, uvod i komentar Hazim Šabanović, (Sarajevo: Svjetlost 1967), 71–94; 101–125; 278–291.

The encounter of Islamic and Christian culture has led to numerous hybrid solutions and a mutual adoption of visual language. After the conquest of Constantinople and the Balkans, the Ottoman-Islamic architecture adopted the Byzantine experience in the design of mosques, while the influence of Islamic art has a significant impact on decorative elements in Eastern Christian visual cultures of the sixteenth and the seventeenth century.[4]

The presence of the Ottoman Empire in the Balkans conditioned the arrival of numerous architects from the empire's provinces. The architect Acem Esir Ali, of Persian ancestry, designed the Gazi Husrev-beg Mosque in Sarajevo in 1530/31.[5] The pre-eminent Ottoman architect Mimar Sinan (1490–1588) also worked in this region.[6] Sinan's structures adorned not only Constantinople, but also the Balkan regions. In Edirne, he built the Selimiye Mosque (1569–1575) for Sultan Selim II. In Trikala, he built the Osman Shah Mosque / Kurshum Mosque (1550–1560). In accordance with the wishes of Grand Vizier Sokollu Mehmed Pasha, Sinan built a bridge on the Drina river in 1577, near Višegrad in Bosnia and Herzegovina.[7] The arrival of the Ottoman architects changed the visual culture of the Balkans; it conditioned its polyphony and hybridity.

In addition to significant cultural contact between Islam and Christianity, the geographical context also occupied an important place in the creation of the Balkan visual culture. The Balkans are a territory that connects the east and west of Europe, as well as the Asian and European space. The Western Balkans' perimeters are on the eastern Adriatic coast, with Renaissance centres such as Šibenik, Ragusa, and Kotor. The eastern perimeters, with the semi-independent principality of Vallachia, bordered the Russian cultural and political sphere. There were the Hungarian and the Habsburg state in the north, while the wide expanses of the Ottoman Empire were in the south. Such geographical positioning has caused the Balkans to become a territory of trade routes, transit, and the encounter of cultures. Numerous artefacts, such as Venetian and Russian icons,

---

4 Verena Han, *Intarzija na području Pećke patrijaršije*, (Novi Sad 1966), 33–108; Andrej Andrejević, *Islamska monumentalna umetnost XVI veka u Jugoslaviji*, (Beograd: Institut za istoriju umetnosti Filozofskog fakulteta – Balkanološki institut SANU 1984); Gülru Necipoĝlu, Anatolia and Ottoman Legacy in: *The Mosque, History, Architectural Development & Regional Diversity*, ed. by M. Frishman and H.-U. Khan, (London 2002), 153–157.

5 Andrej Andrejević, *Islamska monumentalna umetnost XVI veka u Jugoslaviji*, 30–32.

6 Gülru Necipoĝlu, *The Age of Sinan: Architecural Culture in the Ottoman Empire*, (London: Reaction 2005).

7 Godfrey Goodwin, *A History of Ottoman architecture*, London 1971, 313–314; Ömer Turan – Mehmet Z. Ibrahimgil, *Balkanlardaki Türk Mimari Eserlerinden Örnekler*, Ankara 2004, 129.

Persian carpets, calligraphed manuscripts, illustrated and printed books, were all acquired from art centres. Depending on the religious needs and the religious centres of the local population, whose ethnic diversity increased through the immigration of Spanish Jews – Sephardim, after 1492, the acquisition of artefacts was directed towards Constantinople, Moscow, Rome, or Vienna.[8]

The Balkan region was also a region of war. In the border areas the Hungarian, Habsburg, and Ottoman authorities succeeded one another, which conditioned continuous the destruction of existing structures and the construction of new ones. Wars have also conditioned the temporary changes to cultural and visual identity. In 1521, Belgrade came under the Ottoman Empire, and the city was converted from Christianity to Islam. The medieval metropolitan church of Mother of God was transformed to the Sultan Suleiman mosque. After the Habsburg – Ottoman war, Belgrade and northern Serbia were under Austrian rule between 1717 and 1739, which led to a visual shaping of Belgrade, the Serbian capital, as a baroque Catholic-Christian city.[9]

Centuries of coexistence in the Ottoman Empire led to the shaping of a common culture towards the end of the eighteenth and at the beginning of the nineteenth century. This culture is based on an identical visual expression and hybrid intersection of Byzantine and Islamic heritage, with European and Russian practice. This is clearly visible in different aspects – architecture, decoration of public and private buildings, religious art, and especially icon-painting.

The transformation of Ottoman architecture occurs towards the end of the eighteenth century, when baroque-ized and eclectic buildings are erected in Constantinople, as evidenced by the oeuvre of the famous Armenian family of Balyan.[10] The architectural concepts of the so-called baroque-ized public buildings are then encountered throughout the Ottoman Empire, and are also introduced into Christian church architecture in the

---

8 See: Nenad Makuljević, From Ideology to Universal Principles. Art History and the Visual Culture of the Balkans in the Ottoman Empire, in: *Crossing cultures: conflict, migration and convergence*, ed. by Jaynie Anderson, (Miegunyah Press, Melbourne University Publishing, 2009), 98–103.

9 Divna Djurić-Zamolo, *Beograd kao orijentalna varoš pod Turcima 1521–1867*, (Beograd 1977); Pavle Vasić, Barok u Beogradu 1718–1739, in: *Istorija Beograda*, vol. I, (Beograd 1974), 575–583; Nenad Makuljević, Visuality, Conflict and Space. Ottoman, Habsburg and Serbian States (17th-19th Century), in: *Challenge of the Object*, Congress Proceedings – Part 3, Herausgeber G. Ulrich Großmann und Petra Krutisch, Nürnberg 2013, 1126–1128.

10 Alyson Wharton, *The Architects of Ottoman Constantinople: the Balyan Family and the History of Ottoman Architecture*, I. B. Tauris 2015, 5–10; 97–140.

Balkans. The builder's workshop of Andreja Damjanov shows an example of the creation of hybrid architecture, which was executed both for the needs of the Orthodox Church, and for the Ottoman authorities.[11] Certain buildings of Andreja Damjanov, such as the churches in Veles (Macedonia), Niš (Serbia), and Sarajevo (Bosnia and Herzegovina), are among the most representative examples of Balkan architecture. The concept of the above-mentioned buildings did not stem directly from European tradition, but has been created via bringing together different stylistic elements – Byzantine, Renaissance, and Baroque ones. Felix Kanitz, a respected Austrian researcher of the Balkans clearly recognizes this when he talks about the church in Smederevo: "Denn man findet an dieser die Vereinigung des byzantinischen mit dem occidentalen Bauprinzip versucht. Der cincarische Architekt schuf allerdings ein Zwittergeschöpf. Er entlehnte nähmlich die Stirnfassade mit dem Turme den erwähnten ungarische Bauten, den Transept und die Apsis aber die byzantinischen Klosterkirche Manasia. Die Verbindung dieser so ziemlich den Beginn und Verfall der christlichen Kirchenbaukunst bezeichnenden diametralen Bauteile bilden deas verlängerte westliche Schiff und dekorativ einee Masse von Aufputy aus verschiedenstem Material; auch übergipste Eisenornamente kommen vor, die alle Stile von griechischen bis zum Rokoko zeigen."[12]

The decoration of Ottoman-Islamic and Christian public and private buildings is based on identical coloristic and symbolic principles. The concept of decorating mosques and the character of decorations changed towards the end of the eighteenth century, when European decorative models adopted from Mannerist and Baroque culture were introduced.[13] The wall-painting in the "Šarena Džamija" (Colorful Mosque) in Tetovo (Republic of Macedonia), the Bayrakli Mosque in Samokov (Bulgaria), or the Et'hem Bey Mosque in Tirana (Albania) is based on European practice, as demonstrated by the models of painted vases or floral decoration.[14]

---

11 Aleksandar Kadijević, *Jedan vek traženja nacionalnog stila u srpskoj arhitekturi (sredina XIX– sredina XX veka),* Beograd: Gradjevinska knjiga 1997, 14–23; J. Hadžieva – E. Kasapova, *Arhitekt Andreja Dajanov 1813–1878*, Skopje 2001; Nenad Makuljević, Andreja Damjanov: arhitekta poznoosmanskog Balkana, *Zbornik Matice srpske za likovne umetnosti* 38 (2010), 137–149.

12 Felix Kanitz, *Das Königreich Serbien und das Serbenvolk von der Römerzeit bis zur Gegenwart, Dritter Band: Staat und Gesellschaft*, (Leipzig: Verlag von Bernhard Meyer 1914), 786–787.

13 Alyson Wharton, *The Architects of Ottoman Constantinople: the Balyan Family and the History of Ottoman Architecture*, 83–85, 123–140.

14 Ömer Turan – Mehmet Z. Ibrahimgil, *Balkanlardaki Türk Mimari Eserlerinden Örnekler*, 28–29; 204–207; 374–377; Nenad Makuljević, From Ideology to Universal Principles. Art History and the Visual Culture of the Balkans in the Ottoman Empire, 101.

The representative court culture of Prince Miloš Obrenović shows acceptance of common Ottoman – Balkan decorative systems by Christians. Prince Miloš was a semi-autonomous ruler of Serbia (1815–1839), who used Ottoman visual language in his political representation. The wall paintings of his residence in Topčider, near Belgrade, were the same as in public and private Ottoman houses.[15] A similar decorative system was also adopted in Jewish Sephardic art. It is evinced by familiar decoration in synagogues, on ceremonial objects or on illustrated prenuptial agreements – ketubot.[16]

Acceptance of common decorative conceptions is also conditioned by the fact that there were no religious restrictions in the election of master builders and painters for mosques and private buildings. Thus, architects, but also painters and woodcarvers, worked both for Christians and for Muslims. Centuries of living together conditioned the construction of shared visual ideals and the identical attitude to the symbolic understanding of colour. The green was accepted as representative and privileged colour of Islam, so it was rarely used by non-Muslims.[17]

A special form of hybrid culture is seen in icon-painting. The icon is of a dogmatic character in the Orthodox church, which conditioned both their traditionality and continual and large-scale production. Towards the end of the eighteenth and at the beginning of the nineteenth century, a large number of icon painters' workshops were set up in the Balkans. The need for icon painters was conditioned by higher religious freedoms, particularly after the Tanzimat reform of 1839, so, it can be safely said, every Christian church received new painted content. Zographs, from the so-called Debar and Samokov school, excel in the icon-painting workshops.[18] A famous painter, Dimitar Krstevič – Dičo Zograf (1810–1872), can serve as an example of the Balkan zograph.[19] His painting activity encompasses the area of today's republic of Macedonia, Greece, Serbia, Bulgaria, and

15 Katarina Mitrović, *Topčider-dvor kneza Miloša Obrenovića*, (Beograd: Istorijski muzej Srbije 2008), 67–85.

16 *Sephardi Jews in the Ottoman Empire: Aspects of Material Culture*, ed. by Esther Juhasz, (Jerusalem: The Israeli Museum 1990); Nenad Makuljević, Sephardi Jews and the Visual Culture of the Ottoman Balkans, *El Prezente*, vol. 4 (December 2010), 199–212.

17 Nenad Makuljević, From Ideology to Universal Principles. Art History and the Visual Culture of the Balkans in the Ottoman Empire, 100–101.

18 See: Asen Vasiliev, *Blgarski vzroždenski maistori*, (Sofia: Nauka i izkustvo 1965), 151–264; 313–476.

19 On Dičo Zograph see: Asen Vasiliev, *Blgarski vzroždenski maistori*, 179–187; Nenad Makuljević, Liturgija, simbolika i priložništvo: ikonostas Saborne crkve u Vranju, in: *Saborni hram Svete Trojice u Vranju 1858–2008*, (Vranje: Fond "Sveti Prohor Pčinjski" Pravoslavne eparhije vranjske 2008), 52–53.

he was probably educated in today's Albania. Research on Dičo Zograf's icon painting shows that he combined old medieval solutions with iconographic models adopted from Western European experience. Dičo Zograf combined the elements from Byzantine tradition such as the golden colour, reverse perspective, the manner of painting of human figures, with iconographic solutions adopted from early modern European art.

Specific historical and cultural conditions influenced the Balkan territory to acquire a unique visual identity on European soil. The parallel presence of Ottoman-Islamic and Christian visual culture conditioned mutual contacts and transcultural practice. Thus, the visual culture of the Balkans got its hybrid character, conditioned by historical circumstances, cultural contacts, and geographical conditions.

Miroslava Trajkovski

BAF2: DISPOSITIONS

Introduction: Dispositions & Dispositions and Values

1. DISPOSITIONS AND EVIDENCE

Timothy Williamson in "Recognitional Capacities and their Uses" deals with the epistemological significance of recognitional capacities, stressing that "typical recognitional capacities are non-reflective," for "in applying them, we do not use conscious step-by-step reasoning, but our judgment is still evidence-based" (p. 36). Hence, Williamson takes it that "a recognitional capacity for an individual or property X is a capacity to recognize, i.e. come to know, without conscious reflection whether something is or has X on being presented with it in a suitable way under suitable conditions." He gives examples of a chess grandmaster's "recognitional capacity for positions on the board that are a win for black", or how by spending time in Venice, one "can develop a recognitional capacity for paintings by Tintoretto". (p. 36) However, Williamson is particularly interested in their relevance for knowledge-first epistemology's account of evidence and knowledge of moral truths. Relying on the concept of moral recognitional capacities Williamson argues in favor of moral knowledge by *perception*. Finally, he criticizes David Chalmers' method of elimination in the context of verbal disputes ("Verbal disputes", *Philosophical Review*, 120, 2011), claiming that he "has neglected the association of much ordinary vocabulary with recognitional concepts. Eliminating a word typically involves eliminating the recognitional capacity that goes with it, for cats or dogs, or for the subtle or the banal", arguing that "the method of elimination does not leave us with the same cognitive powers, just honed down to fundamentals. Instead, it turns us into idiots, depriving us of most of the distinctions we need to think about the world as adults." (p. 45) Williamson concludes that "during Chalmers' process of iterated elimination, the language is being gradually eroded, bit by bit, in a sorites process. What is left at the end is just a stump. The cognitive power of a natural language consists in large part of a vast array of recognitional capacities associated with the terms of that language". (p. 46)

Slobodan Perović in "The Copernican Principle and Evidence in Modern Cosmology" philosophically explores the concept of the Universe's Physical Predispositions. As he says, "Modern cosmology is not merely an empirical science but a domain deeply embedded in philosophical considerations. Various theoretical models of the universe that are underdetermined by evidence rest upon foundational principles, leading to epistemic and ontological debates about the status of these principles. The Copernican Principle is one example; it has played a crucial role, shaping the conceptual framework of cosmic structure and evolution." (p. 48) According to this principle the universe is homogeneous and isotropic, the questions it raises and Perović approaches are: "Is it a purely operational tool, and if so, how exactly does it "operate" within models? Or does it have unavoidable ontological features and implications indirectly connected to the models' parameters? Does any ontological feature go over and above the operational roles? Is this relevant to a particular model, and if so, how?" (p. 53) Perović sees the need to identify what kind of principle the Copernican Principle is as the immediate philosophical challenge: "Is it an empirical generalization to be tested, or some sort of necessary constraint on cosmological models?" He proposes to answer "these questions by understanding the relationship between *the evidence and the principle*, both in practice and normatively." (p. 52) Perović concludes, "Cosmological models, from the Big Bang to Steady State theories, have reflected competing visions of the universe's nature and origins, while the underdetermination of cosmological theories by evidence and the evolving nature of observational data underscore the provisional status of even the most foundational principles." (p. 58)

2. DISPOSITIONS AND PERSONALITY TRAITS

Aleksandra Pavlović in "Dispositions and brain health: the role of premorbid personality traits in stroke risk assessment|" explores the role of personality traits in stroke recovery and its prevention, for there is "an intriguing association between personality traits (e.g., higher neuroticism and lower conscientiousness) and both Alzheimer's disease and vascular dementia, potentially mediated through shared vascular risk factors and neuropathological changes, despite their distinct etiological origins." (p. 64) According to the five-factor model, basic personality traits include:

extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience.

Pavlović notes that personality changes after traumatic brain injuries have been “widely studied”, but that studies on personality change after stroke are “generally lacking”. (p. 67) Also “Current evidence suggests that higher levels of neuroticism and lower levels of conscientiousness may be novel vascular risk factors, increasing the risk of stroke, vascular brain damage, and other health conditions.” (p. 72) However, Pavlović concludes that “it remains unclear how these findings translate to everyday clinical practice, highlighting the need for additional research. Future studies involving larger and more diverse participant groups with extended prospective follow-ups are essential to better understand the relationship between various personality traits and both brain and overall health. It is also important to investigate other potential confounding factors from both genetic and epigenetic perspectives and to explore potential interventions.” (p. 72)

Goran Knežević in “Major Dispositions in the Psychology of Individual Differences: Conceptualization, Measurement, Origins, and Consequences” deals with dispositions as personality traits, as “stable, broad, cross-culturally universal behavioral tendencies accounting for consistencies in behavior over time and across situations” or “recurrent behavioral, cognitive, or affective tendencies that distinguish one individual from another”, i.e. as individual differences. (cf. p. 80) Knežević explores the history of the idea of personality traits as one natural psychological construct that is often used synonymously with the term dispositions, taken as “enduring tendencies or propensities that guide behavior, emotion, and cognition”. Among types of dispositional constructs in psychology, there are “traits, abilities, aptitudes, attitudes, interests, values, cognitive styles, and more”. (cf. p. 80) One of first systematizations of human personal characteristics comes from ancient philosophy in a famous study, *Characters*, written by Theophrastus. Knežević gives the following criteria for the existence of a trait: descriptive breadth; structural independence; identification/ extraction across methods; identification/ extraction across various groups/ populations; temporal stability; biological basis; predictive validity and relevance.

Though faced with criticisms, “the trait perspective proved resilient.” (p. 82) Particularly important was Five-Factor Model (openness, conscientiousness, extraversion, agreeableness, and neuroticism) which is later accompanied by the HEXACO model, which adds a sixth dimension: honesty-humility, and Knežević agrees about its having a predictive utility in the context of moral behavior, but he and his team propose that “the taxonomy of basic personality trait is to upgrade it with one more, i.e., seventh personality dimension – disintegration/psychoticism – reflecting the proneness to psychotic-like experiences and behaviors." (p. 86) This trait is found to be of crucial relevance for irrational beliefs, e.g., "shown to be a predictor of proneness to COVID-19 conspiratioral beliefs far stronger than socio-demographics". (p. 87)

3. EPISTEMIC DISPOSITIONS

Iris Vidmar Jovanović in "Epistemic Dispositions of Literature: Insights from Literary Tradition and Critical Practice" explores current philosophical debates on the cognitive value of literature. Vidmar Jovanović refutes Gregory Currie's denial of the cognitive value of literature as represented in his *Imagining and Knowing: The Shape of Fiction*, where he claims "that fiction is a source of knowledge and other cognitively valuable states is a dispositional one: what we mean by it is that fiction has the capacity to inspire positive cognitive change in the audience," and "rejects the view that we learn from fiction, suggesting that philosophical belief in epistemic dispositions of literary fiction is misguided and lacks empirical support," and "that we should only accept aesthetic cognitivism if we can in fact prove that readers change in the process or after the process of reading: that they really learn something, that they became more aware of something they did not realize before their exposure to the particular work, that their perspective is wider or better informed, etc." (p. 102) Vidmar Jovanović, in her paper, explores how narrative art in general can be a source of reliable and justified knowledge, at the same time stressing how "just like documentaries about the climate change fail to inspire change in our behavior towards the environment, so too can a novel about the hardships of immigrants fail to inspire a change in how we treat these people." (p. 103) The author deals with the epistemic status of science fiction, some great works of the nineteenth century literature, and some famous mistakes in some well-known classics, pointing out that "not every mistake is relevant in evaluating a work's epistemic dispositions". (p. 116) Vidmar Jovanović acknowledges the "subjective nature of our responses to literature" since "what one picks up from literature depends more on the reader than on the particular work" but she concludes "it is precisely the aspect that makes it so immensely important, helpful and significant in our individual attempts to make sense of ourselves, our experiences and other people." (p. 118)

Ana Kuburić Zotova in "How mindfulness as a disposition impacts epistemic dispositions" examines the impact of a mindfulness disposition-on information processing, and its potential for improving cognitive and emotional capacities. The concept of mindfulness comes from meditative practices. By mindfulness disposition the author means "inclination and inner capacity of paying attention to present-moment experiences with a non-reactive attitude." (p. 123) The author deals with the mindful awareness, seen as a non-reactive attention to the present-moment, exploring its limitations and proposing a model of its integration. Philosophers who particularly recognize the importance of present-moment experience inlcude Kuburić Zotova, James, Husserl, Dilthey, and Gadamer. In general, she notes that "Advocates of mindfulness argue that a changed attitude towards the process of experiencing in the present moment enhances emotional regulation, the ability to control attention, cognitive flexibility, creativity, memory, and the regulation of conation components." (p. 125) The author juxtaposes such views with the findings of those who claim that cognitive processes are controlled and automatic, independent of the reinvestment of attention. So, we come to the notion of meta-awareness, which is a kind of non-judgmental noticing of the contents of consciousness by awareness. (cf. p. 130) That is "mindful meta-awareness insights do not originate from discursive thinking (e.g. deliberation, interpretation, analysis), as reflective insights do, but rather from present awareness." (p. 130) The author concludes that "non-reactive awareness, when trained, is highly beneficial because it promotes dispositions that are foundational and prerequisite for logical-analytical information processing". (p. 140)

Ognjen Milivojević in "Background and Abilities as Dispositions" examines the connection between John Searle's concept of 'background,' and the notion of dispositions. He defines them as follows: background is a set of non-representational, pre-intentional capacities that enable intentional actions, while dispositions are inherent properties that predispose entities to behave in specific ways under certain conditions. (cf. p. 144) Milivojević argues "that since background powers are (fundamental) abilities of the mind and abilities are dispositions, background powers are, therefore, dispositions". (p. 144) His argument has four steps: 1) the analysis and modification of Searle's conception of the background; 2) the analysis of the notion of disposition; 3) the representation of Barbara Vetter's critique of two specific dispositional models of abilities; and 4) the author's argument that ability is a kind of disposition. As to 1), Searle distinguishes between deep and local backgrounds. The former includes universal, basic abilities that are shared by all human beings, and operate in a pre-reflective way, e.g., recognizing objects or walking. The latter is sensitive to social contexts, and includes culturally specific practices, the ability to understand social norms, cultural symbols, and local customs. As to 2), dispositions are contrasted with occurrent (or categorical) properties. As to 3), Vetter denies that abilities are dispositions to do what one intends to do, and that abilities are dispositions to successfully do what one already does. Finally, in step 4), Milivojević, building on Vetter's criticism, proposes a dispositional reading of the background. He concludes "that background powers or, more broadly, abilities are dispositions to behave in a positively adaptive, not purely physiological or physical mechanical manner when opportunities arise." (p. 151)

4. DIPOSITIONS AND ONTOLOGY

Una Popović and Srđan Šarović in "Donkey's dilemma: values or valor?" explore the artist's role in the process of creating artwork. Their thesis is that "the genuine disposition for an artwork to possess value resides in its ontology – that is, in its inner principle". (p. 155) The inner principle is creative, while the external is cultural, both guide the choice between valor and values. While the former is "derived from the work itself", the latter is "shaped by societal constructs". (cf. p. 155) Given this, according to Popović and Šarović, the distinction between valor and value is based on the fact that valor is authentic value. As to values as such, the authors stress that "regardless of which particular value system is adopted, the artist is the one who makes the choice between such value systems. Consequently, he cannot be relieved from the responsibility for his choice". (p. 165) The concept of the 'donkey's dilemma' the authors use to describe the situation of artists who "must decide whether to adhere to the inherent principle of the artwork, ensuring that all stages of its creation are aligned with this internal logic, or to conform to an external principle, thereby adapting the artwork to fit preexisting frameworks." (p. 155) The authors argue that this dilemma is false, for "in either case, the artist remains inextricably bound to the moral and ethical consequences of his choice regarding both the artwork and the principle guiding its creation". (p. 155) One way the authors propose to elucidate the nature of the dilemma is to see it as a choice between *poiesis* (which is ineffable) and discourse (which is an appropriated external principle). They conclude that an artist should take a philosophical stance by constituting and evaluating an ontological model for an artwork.

Dušan Smiljanić in "Dispositions – A Property or A Way of Being" investigates the ontological status of dispositions, asking "Is a disposition some property or a way of being?" This question, he notes, assumes the difference between *ontic* and *ontological*, for it assumes "the difference between a property as an ontic character and a way of being as an ontological character. In both cases, a disposition is viewed as a characteristic of something, and so like something predicable, and not substantial." (p. 167) Specifically, "the sphere of entities and its divisions is called *the ontic sphere*. The sphere of being and its ways is called *the ontological sphere*." (p. 168) In order to answer the initial question, Smiljanić goes into the history of the concept of disposition and ancient disputes on its nature. As to contemporary discussions, he addresses two views: *the linguistic* and *property* views. The former takes it that dispositional terms are not just linguistic ascriptions. The latter sees dispositions as properties. The author discusses three types of dispositions, arguing that "that dispositionality consists of a special way of being of an entity and that it is not a matter of properties or a mere linguistic category." (p. 181) Finallly, Smiljanić compares Aristotle's and Heidegger's concepts of disposition. He concludes that these two views are comprehensive concerning the issues of the ontological status of dispositions, so that after Aristotle and Heidegger, "the approach to dispositions must be focused on the research of concrete phenomena, primarily in the human world of spirituality, culture, politics, etc." (p. 185)

BAF2+: DISPOSITONS AND VALUES

5. DISPOSITIONS AND VIRTUES

Amber Riaz in "Moral Learning for the Wretched of the Earth" analyzes two individuals raised in different moral environments that opposed to one another. Still, Riaz points out that the difference need not necessarily be in abstract moral principles, but in their application. She notes "that on a standard account of knowledge of concepts and principles on which it is a priori, concepts and principles do not come with a manual on how to apply them. Instead, how to apply them are additional and important skills that have to be learned in experience, where the notion of experience is to be understood very broadly as involving direct real life experiences as well as reading different genres of literature and non-literary writings, viewing artwork, watching films, theatre and so on." (p. 193) She compares the behavior of an agent coming from a defective moral environment with his "acqueried" behavior when he is put in a healthy environment. She claims that such agents develop conflicting intuitions, resulting in epistemic self-doubt that "constitutes a kind of moral learning for such agents; they also play a significant role in developing moral understanding". (p. 191) Riaz coins a term WOTE (Wretched Of The Earth) for epistemically non-ideal and morally unlucky agents, arguing that an agent's realization that s/he is a WOTE is epistemically valuable for his/her moral learning. (cf. p. 198) She concludes that WOTEs "fortunate enough to realize that their moral environment is defective, can make good use of their epistemic self-doubt and internal tension generated by conflicting intuitions to marshal cognitive resources other than affect, and capitalize on the fragmentation of his mind, to gain moral understanding". (p. 203)

Marcin Trepczyński in "Virtues as Dispositions: Different Approaches in Medieval Analytic Thought" deals with the question whether "hexis" (ἕξις) in the definitions of virtue is a kind of disposition. He also examines the relationship between habitus and virtue. The author uses "*habitus*" and not "habit", as current usage of the latter deviates from its historical usage, which is better captured by the term "disposition". The author cites the following definition of virtue from the twelfth century: "virtue is the condition or disposition of the well-ordered mind" (*virtus est habitus mentis bene constitutae*). In one translation, Aristotle's definition of virtue reads "Virtue then is a settled disposition of the mind determining the choice of actions and emotions... (...)", so as Trepczyński stresses "ἕξις is a disposition, however, a special kind of disposition, settled in a humans, which corresponds to the idea of stability. And consequently, virtue is a kind of disposition as well". (p. 209) The author continues with Abelard and his disciple, John of Salisbury, elucidating the Aristotelian distinction between διάθεσις (*dispositio*) and ἕξις (*habitus*). Trepczyński devotes particular attention to Stephen Langton, who "reads *habitus* as an ability or capacity of a person", (p. 215) and "provides us with insights concerning the concept of *habitus* as compared to the notion of disposition". (p. 216) One such insight is quoted, it reads: "that to be virginity is for virginity something accidental, similarly, for a disposition to be a disposition, because it will be a disposition when it will not be a disposition, but a *habitus*." (p. 216) Turning to Thomas Aquinas, the author explains changes in thinking about virtues, dispositions and *habitus* that he introduced by being the first to call *habitus* a disposition (*dispositio*). (p. 220) The author concludes that "according to Thomas Aquinas: virtue is disposition", (p. 221) also "that the Greek 'hexis' used in the Aristotelian definition of virtue is a kind of disposition". (p. 223)

6. DISPOSITIONS AND EMOTIONS

Damir Smiljanić in "Cool Philosophy or the Art of Restraining from Judging without being Indifferent" deals with the question about the nature of a philosophical restraint from judgment, asking if such restraint is based on some special (personal) disposition or *value attitude* accepted by a cognitive subject. The author argues that *indifference* is not a suitable descriptor, unlike *coolness* and *prudence*. He starts from the distinction between *dispositions* and values and characterizes indifference as a *non-evaluative disposition*. Given this, Smiljanić investigates whether indifference can be a *manifestation* of a *value attitude*. He distinguishes *psychological* and *philosophical* descriptions of indiference: indifference as a *character trait* or indifference as a philosophical or ethical *attitude* (i.e. *value attitude*). The author in his analysis of restraint from philosophical judgment considers the folowing three forms: "1) restraining from judgment in the ancient tradition of *radical scepticism* (*Pyrrhonism*) (the so-called *epoché*), 2) observing a problematic situation through the eyes of an unbiased observer (as suggested by Adam Smith), 3) Max Weber's recommendation that those dealing with science should – at least at the academic level – be *unprejudiced*." (p. 230) Smiljanić claims that "the sceptics' radical suspension of judgment, Smith's concept of the impartial spectator, and Weber's appeal for value neutrality in science – demonstrate that, in a philosophical context, indifference is indeed a positive phenomenon". (p. 237) Moreover, the author adds, indiference is an ethically desirable habitus, it holds a certain value, it is even "a necessary condition for 'effective' philosophizing". Finally, Smiljanić concludes that to keep a cool head in philosophy means to resist "making hasty judgments about intensely debated matters". (p. 239)

Isidora Novaković in "Dispositions to Tragic Emotions" asks if there are dispositions to tragic emotions, and if there are, what evokes them, is it something in the tragedy itself? She starts from Aristotle's definition of tragedy; namely, his understanding of tragic suffering and tragic emotions (fear and pity) and his general understanding of emotions as a crucial for moral virtues. According to Aristotle "tragedy portrays people in action, aiming to present us with characters on stage people who can be better, worse or similar to us – characters who, themselves will have certain virtues and flaws". (p. 242) In tragedy, people are involved in "events that evoke fear and pity. These effects occur above all when things come about contrary to expectation but because of one another." (p. 243) Fear and pity are evoked under certain conditions: pity is felt for a person undeserving of his misfortune, fear for a person similar to us. (p. 248) In general, for Aristotle "what we fear for ourselves excites our pity when it happens to others." (p. 246) Novaković analyses the above ideas through the examples of Greek tragedies. She explores Aristotle's view on the relationship between emotions, dispositions and virtue. She uses the term 'disposition' for what Aristotle calls capacity, while 'habit' is used for what

Aristotle calls disposition, taking it that dispositions are the bridges connecting emotions with habits. (p. 249) As an illustration, Novaković says: "we can imagine an emotion as a person walking down the road (disposition), in order to arrive at the final destination (habit)". (p. 249) The author also deals with the notion of *hamartia* (a fatal mistake of the tragic hero that leads to tragedy), concluding that "the goal of tragedy is to evoke tragic emotions that, in turn, lead to catharsis". (p. 256)

7. DISPOSITIONS IN AESTHETICS AND ETHICS

Aleksandar Risteski in "The Anthropological Implications of Peirce's Aesthetics" starts from the fact that Perce's work on aesthetics was only fragmentary. Still, he aims to reveal the importance of Peirce's aesthetics for his general philosophy and its place in his phaneroscopy. The author notes that "the implications for phaneroscopy are that it demands the similar kind of attention ("observational power"), or noticing of all the elements that can be related to an observed phenomenon, irrespective of its ontological status, the elements that usually can be overlooked because we assume how we should observe the phenomenon." (p. 267) Also, an aesthetical quality is related to the category of Firstness, while "aesthetics studies that which is admirable and desirable in itself". (p. 269) Risteski compares Perce's ideas with those of Plato and Kant, noting that Peirce did not accept traditional views of beauty, he "argued that harmony and symmetry are only conditions for beauty to appear, and not something identical to it". (p. 269) The author concludes: "Aesthetics should observe the mechanics of forming and following an ideal. A man, as a teleological system, cannot do without having different ideals. An ideal is that which allows our behavior to be described as purposeful and goal-oriented. Aesthetics recognizes the close relationship between the concepts of the highest good, and aesthetical qualities and experiences. That is why an account on the teleological form of our conduct would require a theory like that. Ultimately, such a theory would provide a strong aid to a philosophic-anthropological account on what a man is." (p. 277)

Natalia Tomashpolskaia in "Ludwig Wittgenstein's critique of the dispositional theory of values" deals with Wittgenstein's critique of the dispositional theory of values, arguing that it poses a challenge to both naturalistic and relativistic tendencies. The paper explores Wittgenstein's rejection of the idea that values can be reduced to subjective or cultural preferences, his understanding of *as-if* conventions, his distinction between aesthetics and psychology, his claim that ethics is fundamentally distinct from moral customs or sociological descriptions, and his distinction between taste and aesthetic judgement. The author explores Wittgenstein's distinction between the rules of *representation* and *conventions* which "brings us to the notion of the 'mystical', which lies beyond any such conventions", she argues "that the later Wittgenstein did not entirely dismiss all 'mystical' views. In his later notes, remarks, and, especially, in his conversations with friends, we find intriguing passages on ethics, aesthetics, religion, and consciousness that do not neatly fit into the framework of Wittgenstein as merely an analytic philosopher or a conventionalist." (p. 288) Tomashpolskaia concludes that "justification ultimately depends on one's own reactions – specifically, feelings of admiration or disgust. Without such personal responses, ethical propositions lack the weight or authority that might make them meaningful. In this way, Wittgenstein emphasises the *subjective* foundation of ethics, the significance of an ethical proposition is rooted not in external facts but in individual emotional engagement." (p. 292)

BALKAN
ANALYTIC
FORUM

BAF2: Dispositions

1. DISPOSITIONS AND EVIDENCE

https://doi.org/10.18485/baf.2025.2.ch1

Timothy Williamson

RECOGNITIONAL CAPACITIES AND THEIR USES

Abstract: A recognitional capacity for a kind is a capacity to come to know whether an appropriately presented object is a member of that kind; similarly, there are recognitional capacities for properties and for particular individuals. Recognitional capacities are typically non-reflective and fallible. Although often associated with words of a natural language that refer to the kind, property, or individual, they are not part of the semantics of those words and are not required for linguistic competence with them. Recognitional capacities for properties expressed by moral terms are argued to play a major role in moral epistemology, and even in moral perception. Arguably, David Chalmers' recently proposed methodology for identifying verbal disputes and identifying bedrock concepts depends on neglecting the epistemic role of recognitional capacities associated with ordinary words.

Keywords: recognitional capacities, moral epistemology, moral perception, verbal disputes, David Chalmers.

In this paper, I will briefly explain what recognitional capacities are and how we humans and other animals use them, and derivatively how we philosophers can use the category of recognitional capacities to better understand everyday phenomena that they have managed to be puzzled by.

To illustrate: a recognitional capacity for horses is a capacity, when you see something, to recognize whether it is a horse. Of course, such capacities are never perfect. You may see a horse in the distance, or in bad light, and mistake it for a cow, or be uncertain whether it is a horse or a cow. Likewise, you may see a cow in the distance, or in bad light, and mistake it for a horse, or be uncertain whether it is a horse or a cow. Most people who understand the word 'horse', or a synonymous word in another language, have an associated recognitional capacity for horses, though that is not strictly needed for understanding the word. You could understand it by having heard others talk about 'horses', without ever having seen a horse, or even a picture of one.

Not all recognitional capacities are visual. A blind musician may have an auditory recognitional capacity for oboes. She has a capacity, when she hears something, to recognize whether it is an oboe.

One can also have a recognitional capacity for a particular individual. Many people have a capacity, when they see something, to recognize whether it is Donald Trump (or a picture of him), though such a capacity is not needed for understanding the name 'Donald Trump'.

Recognitional capacities are typically *unreflective*. When you recognize a horse, or Donald Trump (perhaps in a picture), you do not normally engage in conscious calculation or inference. You just judge 'It's a horse' or 'It's Donald Trump'. Of course, complex brain processes are at work behind the scenes, but they are not suited to appearing on the stage of consciousness.

I take the phrase 'recognitional capacity' from the work of Michael Dummett. However, I do not expect recognitional capacities to do the sort of work he introduced them to do. Dummett introduced recognitional capacities as candidates to be the more or less Fregean senses of terms that could not plausibly be treated as abbreviating verbal descriptions, at least for a given speaker (Dummett 1973). By contrast, I do not cast recognitional capacities in any strictly *semantic* role, although they may play a major *meta*semantic role in helping fix the referents of the associated terms, since an associated recognitional capacity can be central to how speakers use their terms, what they apply them to and what they do not. Arguably, dispensing recognitional capacities from any semantic duties enables us to give a more realistic account of their epistemology, free from ill-motivated and unrealistic constraints.

In the next section, I will explain how acknowledging recognitional capacities for moral properties enables us to give a more psychologically plausible account of moral epistemology. That account draws on a more general defence of moral realism that I have developed elsewhere. In the final section, I will explain how, by reflection on the dependence of our use of natural language on recognitional capacities, we can diagnose what is wrong with a widely held view of concepts, one which in particular underlies David Chalmers' discussion of verbal disputes (Chalmers 2011).

Other philosophical uses of the category of recognitional capacities include cases where acquiring non-deductive evidence for a conclusion triggers a recognitional capacity by which one comes to know that very conclusion, thereby making it part of one's evidence base; acknowledging such double-takes helps fit inferential evidence into the framework of knowledge-first epistemology (Williamson 2023, in response to Dunn 2014). Those more technical applications will not be discussed in this paper.

Moral recognitional capacities

One form of attack on moral knowledge consists in challenging moral realists to explain *how* they know moral truths. Their opponents suspect that sooner or later moral realists will have to fall back on a suspiciously convenient faculty of *moral intuition*.

An example: I judge 'Poisoning Alexei Navalny was wrong'. Is that a moral intuition? I know that Alexei Navalny was poisoned, from various reliable news sources; it is a normal case of knowledge by testimony. I then make the moral judgment that poisoning him was wrong. Although I did not reach my conclusion by conscious step-by-step reasoning, I can provide non-deductive support for it, by citing known circumstances of the case and other considerations. In those respects, it is not so different from my non-moral judgment 'Poisoning Alexei Navalny was premeditated'. In that case too, although I do not reach my conclusion by conscious step-by-step reasoning, I can provide non-deductive support for it, by citing known circumstances of the case and other considerations. In each case, I applied a new term ('wrong', 'premeditated') in the light of my evidence. Presumably, I *can* thereby come to know that poisoning him was premeditated. If my judgment 'Poisoning him was premeditated' thereby depends on 'intuition', presumably so too does my judgment 'Poisoning him was wrong', but since one can achieve knowledge by depending on intuition, as in the other case, there is no obvious cause for alarm.

More generally, we have *recognitional capacities* for properties and kinds of many types, enabling us to recognize whether a given case instantiates them. Most people have recognitional capacities for many species of plants and animals and for many types of artefact—types of clothing, types of furniture, types of household utensil, types of vehicle, types of building, and so on. They can recognize various types of weather, various types of art and music, and various types of behaviour: hasty or leisurely, careless or careful, confident or timid, rude or polite, cold or warm, serious or humorous, hostile or friendly, cruel or kind. Non-human animals too have recognitional capacities for various kinds of animal—their own species, predators, prey, males, females, young, mature—and for various kinds of food, building material for nests or dams, and so on.

Linguistic competence with an associated term is not necessary for having a recognitional capacity. A species of birds that nested only in elm trees would have a recognitional capacity for elms. Nor is linguistic competence with the term sufficient for having the recognitional capacity: by ordinary standards, Hilary Putnam was linguistically competent with the word 'elm', but by his own admission he had no recognitional capacity for elms. However, a recognitional capacity associated by a core of users of a natural kind term may play a crucial role in fixing its reference to the natural kind (Brown 1993).

Although many recognitional capacities are fitness-enhancing from an evolutionary perspective, evolutionary constraints put no clear upper bound on what recognitional capacities we can acquire. A chess grandmaster has a recognitional capacity for positions on the board that are a win for black; spending time in Venice, you can develop a recognitional capacity for paintings by Tintoretto. No sound evolutionary argument excludes such recognitional capacities, despite the tenuous relation between evolutionary pressures and chess or art history. Similarly, to expect some evolutionary debunking argument to exclude recognitional capacities for moral properties would betray a remarkably superficial understanding of evolutionary theory. In particular, it is arguable that the properties expressed by moral predicates are not of some radically different metaphysical kind from those expressed by non-moral predicates such as 'is a win for black' or 'is by Tintoretto' (Williamson 2025, chapter 1).

As the examples above illustrate, typical recognitional capacities are non-reflective. In applying them, we do not use conscious step-by-step reasoning, but our judgment is still evidence-based. To describe recognitional capacities in all those cases as based on a faculty of 'intuition' merely obfuscates what sort of pattern recognition is going on.

A different view of particular moral judgments like 'Poisoning Alexei Navalny was wrong' is widespread in moral epistemology; its recent advocates include Paul Boghossian and Christopher Peacocke. On such a view, such judgments are *inferential*. They come from arguments like this, where 'D' is schematic for a qualitative *non-moral* description:

Premise 1 Poisoning Alexei Navalni was D.

Premise 2 Whatever is D is wrong.

Conclusion Poisoning Alexei Navalny was wrong.

The argument is deductively valid, with Premise 2 read as a universal generalization over all times. Premise 1 is particular, purely non-moral, and *a posteriori*; Premise 2 is general, moral, and supposedly *a priori*; the Conclusion is particular, moral, and *a posteriori*. On this view, the moral element comes from a capacity to assess general moral principles *a priori*. By contrast, on the view defended in this paper, the moral element comes from a recognitional capacity to assess particular cases morally *a posteriori*.[1]

The inferential view of moral judgment is implausible, both psychologically and epistemologically. Psychologically, we are not aware of making any such deduction, or even of entertaining any such universal generalization as Premise 2; nor have we any evidence that what goes on in us unconsciously takes any such artificial form. Epistemologically, we are typically *more* confident of the Conclusion than of Premise 2, because doubts as to whether 'D' excludes all potential counterexamples absent from the Navalny case pose a threat to Premise 2 but not to the Conclusion. Thus, even if some of our confidence in the Conclusion comes from Premise 2, not all of it does. Formulating universal moral principles like Premise 2 takes us into a realm of proto-philosophical speculation that we did not enter merely by making a moral judgment on a particular case.

The envisaged inferential alternative fails for a deeper reason too. For where does Premise 2 itself come from? If we are in a position to have ordinary pre-theoretic knowledge of Premise 1, the non-moral description 'D' will be couched in reasonably accessible, not too complicated terms. But such a description is likely to make Premise 2 exception-prone and so false, even in our own eyes—poisoning Hitler in 1939 would not have been wrong, and so on. Premise 2 is hardly innate, nor were we taught it as children.

Will moral realists invoke 'intuition' here, to explain our putative knowledge of Premise 2? That is what their opponents suspect. What is most dodgy about such reliance on 'intuition' is not that it sounds spooky but that it fails to engage with the semantic structure of the sentence. Premise 2 is a universal generalization of the form 'Whatever is F is G'. Normally, to recognize such a sentence as true, without matching it to a list of pre-given truths, one must recognize some necessary or contingent connection between the two properties, being F and being G (here, being D and being wrong). One can articulate that connection as linking the simpler sentences '*x* is F' and '*x* is G', with the variable '*x*' for an unspecified instance, prior to introducing the universal quantifier ('whatever'). More specifically, by the informal analogue in natural language of a standard introduction rule for the universal quantifier in a system of natural deduction, having verified '*x* is G' conditionally on the supposition '*x* is F', one can then verify the universal generalization 'Whatever is F is G'. That is logically the most basic way of verifying a universal generalization

1 For an account of such a form see Peacocke 2004: 198–231 and Boghossian's discussion of moral intuition in Boghossian and Williamson 2020. For other contemporary invocations of moral intuition see Huemer 2005 and Wedgwood 2006.

of that form. To verify Premise 2, it means supposing an unspecified instance '*x* is D' and verifying '*x* is wrong' under that supposition. But '*x* is wrong' is just an unspecified instance of the same pattern as the Conclusion, 'Poisoning Andrei Navalny was wrong'. Moreover, making a judgment under a supposition, imaginatively, is just the offline analogue of making the judgment unconditionally, 'live' online, in a particular case. To make such a judgment we need just the sort of recognitional capacity (still for 'wrong') invoked by the simpler account above.[2]

In brief: assenting to a sentence on the basis of intuition was not intended to bypass semantic understanding of that sentence, but instead to work through it; the most natural way of applying that approach to Premise 2 makes the inferential apparatus redundant.

Might our assent to Premise 2 come from some more roundabout process of reasoning? One could postulate that we derived Premise 2 deductively from more basic moral principles, but that only postpones the general problem to those latter principles, on pain of an infinite regress: how do we know the more basic moral principles? Similarly, to postulate that we arrived at Premise 2 inductively or abductively returns us to the original problem, since the inductive or abductive evidence for Premise 2 would presumably itself include further particular moral claims like the Conclusion.

We could try replacing the universal generalization 'Whatever is D is wrong' as Premise 2 by a merely generic generalization or *ceteris paribus* 'law', 'D actions are wrong', making the argument non-deductive. But that modification does not address the underlying problems, which arise in similar ways for the new Premise 2. Without the recognitional capacity for particular cases, we still cannot verify the generic generalization.[3]

The upshot of these considerations is that inferential accounts fail to present a genuine alternative, since they will not work without a recognitional capacity, the need for which they merely obscure. Their extra complexity brings no extra explanatory power.[4]

---

2 For the suppositional assessment of universal generalizations see Williamson 2020: 142–6. For the connection between offline and online judgments see Boghossian and Williamson 2020: 121–2 and 179–81. Many of the epistemological and psychological misconceptions in the literature on 'moral intuitions' are analogous to those in the metaphilosophical literature on 'philosophical intuitions'; I discuss them at length in Williamson 2021.

3 The account of the suppositional assessment of generalizations in Williamson 2020 is applied to generics.

4 Overestimates of the role general principles play in ordinary moral cognition can also have distorting effects on semantic theories of moral discourse. For example, according to Alex Silk's Discourse Contextualism, 'Moral uses of deontic modals

Of course, recognitional capacities themselves involve hidden complexity, like the neural processes underlying face recognition. Any recognitional capacity for a moral property, however limited, will involve such hidden complexity too. But the key objection to the inferential account is *not* that it postulates hidden complexity. Rather, it is that the inferential account is *regressive*. For it analyses the cognitive process by which we make a singular ascription of 'wrong' into an inference whose major premise (Premise 2) is a generalized ascription of 'wrong' (typically a novel one), our assent to which itself depends on a prior singular ascription of 'wrong', albeit offline. If every singular ascription of 'wrong' depended on a prior singular ascription of 'wrong', we could never get started.

By contrast, applying a recognitional capacity involves no such regress. No component of the neural processes underlying face recognition is itself another recognitional process as complex as the one we started with (a homunculus recognizing a face on the screen of a mental cinema). The same applies to our recognitional capacities for moral properties.

Once we have a (limited) recognitional capacity for 'wrong', we can use it to start formulating general moral principles and testing them against particular cases, real and imagined. We may adopt some of those principles, and even start using them to correct the deliverances of the original capacity, where we detect a bias or anomaly. We may indeed make some singular ascriptions of 'wrong' inferentially, using those principles as premises, as with the argument from Premise 1 and Premise 2 to the Conclusion. That would not be regressive. What triggers the regress is the assumption that *every* singular ascription of 'wrong' is inferential.

That we have recognitional capacities for moral properties is no mere postulate of moral realists. Much ordinary moral discourse takes it for granted, for example when we describe someone as recognizing that an action was wrong. Phenomenologically, much ordinary moral judgment (and misjudgement) feels like the exercise of recognitional capacities for moral properties.

[...] presuppose a body of moral norms endorsed in the context' (2017: 231); a discourse-level parameter in the semantics 'represents norms accepted for purposes of the conversation' (ibid: 228). These general norms are needed because 'The truth-conditional contents of deontic modal sentences are propositions about logical relations (e.g. implication, computability) between propositions and premise sets' (ibid: 226). Silk takes such general norms to be shared (for conversational purposes) even when speakers have moral disagreement about particular cases. The picture of everyday moral disagreement as fundamentally about logic is a philosopher's dream. Although one can probably think up some general moral claim or other so bland that all parties to the dispute would assent to it, that does not mean that its sparse logical relations were in dispute. There is no good evidence for the normal conversational presence of such general moral principles.

Naturally, a philosopher can still insist that what we have are *failed* recognitional capacities. For a moral anti-realist, there are no moral properties to be recognized. For an error theorist, the would-be recognitional capacity is like one for the property of being bewitched. But such views are irrelevant to the present dialectic. By hypothesis, there are moral truths, for the question is whether a distinctive problem of moral knowledge still arises under that supposition. It is a short step from moral truths to moral properties within a standard semantic framework, since moral truths require moral predicates to express moral properties in the relevant sense.[5] Thus we may assume that there are moral properties like being wrong for the associated recognitional capacities to track.

Of course, not all recognitional capacities are good at their job. Someone could hold that, although the word 'wrong' does express the moral property of being wrong, the associated recognitional capacity is bad at tracking that property, indeed, is hopelessly unreliable. Some moral realists take our would-be recognitional capacities for moral properties to have been definitively superseded by the clear principles of their favoured moral theory—Kantianism, utilitarianism, whatever. However, such views pose little threat to the present strategy. They hold that one can do much better epistemologically by theoretical reflection on moral principles than by reliance on common sense recognitional capacities. Indeed, their scepticism about the latter is justified by the alleged epistemological superiority of the former when the two conflict. The suggestion is not that the moral domain is epistemically inaccessible, just that it is better accessed by theoretical reflection than by ordinary recognitional capacities. One may doubt that the theoretical reflections at issue are as cogent as claimed, but there is no need to press such doubts here, for either way the moral domain is epistemically accessible.

An alternative kind of argument for the unreliability of moral recognitional capacities invokes individual and cultural variation in their deliverances. Such reasoning grants the associated moral properties, at least

5 In epistemological contexts, we often need to make finer-grained distinctions amongst mental states than can be made simply by ascribing propositional attitudes such as knowledge and belief to coarse-grained intensions. We can do so by relativizing propositional attitudes to the guise under which the thinker entertains the proposition. For example, the guise may be a sentence of the speaker's language, perhaps in a particular conversational and perceptual context. You can accept the same proposition under one guise without accepting it under another. In order to avoid bogus counterexamples, when cases are described in terms of propositional attitudes, care must be taken to distinguish guise-sensitive from guise-insensitive ascriptions; natural languages can be very misleading in this respect. See Williamson 2024a for detailed discussion and defence of the intensionalist approach. Anyway, none of this affects the connection between truth and properties.

for the sake of argument, and does not privilege some alternative form of epistemic access to the moral domain. The idea is just that if one person or group judges an action as wrong, while another judges it as right, one side is in error (since we are assuming moral realism, no relativistic doublethink is in play). Thus, the extent of moral disagreement puts an upper limit on moral reliability.

Such arguments from disagreement are not illegitimate in principle, but must be applied with care. In particular, we must remember that moral knowledge, like non-moral knowledge, requires only local reliability. For example, suppose that members of society S judge that actions of type A are wrong, while members of society S* judge that actions of type A are right. If members of each society restrict their judgment to actions performed within their own society, so far there is no disagreement, since whether an action of a given type is right or wrong may depend on the social setting: a physical movement may constitute an offensive gesture in S but not in S*. However, suppose that members of each society generalize to the other: members of S judge that instances of A are wrong, whatever society they are performed in, while members of S* judge that instances of A are right, whatever society they are performed in. Now there is genuine disagreement, since the judgments are made about the same domain of actions. At least one side is in error somewhere. Nevertheless, it may still be that members of each society are reliable about whether A-actions performed in their own society are wrong: instances of A performed in S are wrong, while instances of A performed in S* are not wrong. We should not be surprised when people are better at understanding the moral significance of actions performed in their own society than of actions performed in another society, with which they are less familiar. Thus, members of S may *know*, of particular A-actions performed in S, that they are wrong, while members of S* *know*, of particular A-actions performed in S*, that they are right.

That example is obviously very simple and schematic. Sometimes, the asymmetry may run the opposite way, with members of each society blind to its own faults but alert to the faults of its neighbours. The point is just that widespread moral disagreement is quite compatible with widespread moral knowledge of specific matters.

Gestures at the extent of moral disagreement are often quite perfunctory. In surveying the data, one danger to avoid is *cherry-picking*. For example, given any moral belief, one can probably find people somewhere with a contrary moral belief. But that shows nothing special about morality. Human individuals and human societies are very various. Given any *non-moral* belief, one can probably find people somewhere with a contrary non-moral belief. Disputes between proponents and opponents of the theory of evolution are no more likely to be resolved than disputes between proponents and opponents of a right to abortion. On both moral and non-moral matters, disagreement is typically louder and more attention-grabbing than agreement, making us liable to underestimate the latter's prevalence.

Still, there *is* plenty of moral disagreement. Some philosophers may therefore be tempted to dismiss local patches of correlation between moral belief and moral truth as products of chance rather than cases of moral knowledge. Wouldn't one expect such patches, if moral belief and moral truth varied independently of each other over a large enough domain? As just explained, that attitude may well underestimate the extent of moral agreement. But it also ignores another crucial factor: the *metasemantics* of moral terms. Like other words, 'wrong' does not get its intension by magic. Speakers are willing to apply it in some cases and not in others; the patterns of their use help determine its reference. A principle of charity constrains correct interpretation. Arguably, the appropriate general principle maximizes the *knowledge* expressed by the relevant discourse: for 'right' and 'wrong' in their moral senses, that is moral knowledge.[6] Although the correct interpretation is the resultant of several forces, correlations between moral belief and moral truth are far from random: they manifest in part a constitutive pressure towards knowledge.

Recognizing moral recognitional capacities also help us make sense of a category that has proved controversial: moral knowledge by *perception*. It has sometimes been regarded as a case where moral epistemology diverges from non-moral epistemology. I will suggest that it is much less exceptional, and much less problematic, than it may sound.

Perceptual language is often used in an extended or metaphorical sense applicable to both sensory and non-sensory cognition, including moral cognition. Someone can *perceive* a research programme as flourishing, or come to *see* that it is degenerating. Likewise, someone can *perceive* slavery as acceptable, or come to *see* that it is wrong. Such examples are no more contentious than moral knowledge; leave them aside. The question is whether there can be moral knowledge by genuinely sensory perception. For example, can one gain moral knowledge *by seeing*, in a sense which requires one to use one's eyes, though of course also one's brain?

Seeing-that must be distinguished from object-seeing. One can see an Anglo-Saxon coin without seeing *that* it is an Anglo-Saxon coin, but

6 For the role of charity in interpretation see Grandy 1973 and Davidson 1977. For a knowledge-maximizing principle of charity see Williamson 2007: 248–78. For a reinterpretation of that principle in terms of an account of mindreading in humans and other animals, see Williamson 2024b.

an expert numismatist can see *that* it is an Anglo-Saxon coin. If we have an encapsulated visual module, impervious to background information, it will not by itself generate judgments like 'That's an Anglo-Saxon coin'. Nor, presumably, will it generate judgments like 'That's a giraffe' or 'That's Novak Djoković'. Perhaps it confines itself to shapes and colours. Indeed, by itself, it will generate very little of the seeing-that which we need in science and in ordinary life—potentially including the ordinary moral life.

In the moral case, the issue is not whether one can see special moral objects, but whether one can see *that* an object has a moral property. Sometimes, one can see *that* a kicking was deliberate. Sometimes, one can also see *that* it was wrong. Some philosophers may insist that one only sees that the kicking had various non-moral properties, and then infers that it was wrong by an argument like that above from Premises 1 and 2 to the Conclusion. However, there is no good reason to impose such an inferential structure. Seeing-that is not restricted to contents internal to a primitive module for vision. Sometimes, an expert chess-player sees that a position is a win for black, and an expert boxer sees that his opponent is tiring. By other cues, sometimes a psychologically competent person sees that a kicking was deliberate, and a morally competent person sees that it was wrong.

Seeing-that is a way of knowing-that: when you see that something is so, you thereby know that it is so (Williamson 2000: 33–41). Thus, when you saw that the kicking was wrong, you thereby knew that it was wrong. Consequently, there is moral knowledge by perception.[7]

In those formulations, the words 'way', 'thereby', and 'by' must not be read instrumentally. Seeing that something is so is not a *means* to knowing that it is so. Rather, seeing that it is so is *already* knowing that it is so; seeing-that is a subtype of knowing-that. Seeing that the kicking was wrong is a form of knowing that it was wrong. One's recognitional capacity for wrongness is triggered visually, but it can also be triggered in other ways, for example by hearing someone describe the kicking in words.

'She saw that the kicking was wrong' can be correctly paraphrased as 'She saw the wrongness of the kicking'. The trouble is that someone may then misinterpret the paraphrase by misreading 'saw' in terms of object-seeing in place of seeing-that, casting the wrongness of the kicking as some strange kind of visible object. Such confusions can easily result in the impression that moral perception would be weird. But the confusion is not confined to the moral. Perhaps, in another case, she saw that the kicking was unintentional. 'She saw that the kicking was unintentional' can

7 See McGrath 2004 and 2019 for more on more moral knowledge by perception.

be correctly paraphrased as 'She saw the unintentionality of the kicking'. Someone may then misinterpret the paraphrase by again misreading 'saw' in terms of object-seeing in place of seeing-that, casting the unintentionality of the kicking as another strange kind of visible object.

Bad theories of perception can be used to make moral perception look implausible. Fortunately, the badness of those theories emerges in non-moral perception too. Moral perception is unexceptional.

Recognitional capacities and graceful degradation

On a picture associated with the traditional programme of conceptual analysis, most concepts are complex and built up from simple concepts, into which they can be analysed. David Chalmers has offered a sophisticated defence of a similar view, for example in his work on verbal disputes (Chalmers 2011). He proposes to distinguish between verbal and non-verbal disputes by the method of *elimination*: if a word is used in a dispute, try eliminating that word from your vocabulary and see whether you can still have something like that dispute. If the dispute survives, so provisionally does the hypothesis that it is non-verbal: at least, it did not depend on the word at issue. But if eliminating the word eliminates the dispute, that usually indicates that the dispute was merely verbal, since it depended on the word. However, Chalmers allows for exceptions, cases of 'vocabulary exhaustion', when the word expresses what he calls a 'bedrock concept', and eliminating it reduces the overall expressive power of the language. For example, he suggests, the words 'conscious' and 'ought' may express bedrock concepts. But he insists that such cases are rare and exceptional. On his view, one can progressively eliminate word after word from the language without reducing its overall expressive power, until one is left with a tiny non-redundant core of exceptional words expressing bedrock concepts, from which all the others can be reconstructed, more or less. That core has the same expressive power as the original language.

How well does Chalmers' picture fit natural languages? Suppose that you and I see an animal in the distance, which quickly disappears round a corner. I judge: 'It's a cat'; you judge 'It's a dog'. We are talking about the same animal. It is a prototype of a non-verbal dispute. But what happens if we eliminate the word 'cat', or the word 'dog', or both, from our vocabulary? We might happen to have the words 'feline' and 'canine' in our language, with the same meanings as 'cat' and 'dog' respectively, on which we could fall back to rephrase our dispute, though they are not needed for the original dispute, and in any case the same problem arises for the dispute between 'It's a feline' and 'It's a canine' once the words 'cat' and 'dog' have been eliminated. Normal speakers of a natural language do not have descriptive equivalents of ordinary natural kind terms in 'more basic' terms available to them, as Kripke and Putnam made clear in the 1970s. Of course, if we had pictures of a prototypical cat and a prototypical dog, we could point at them and argue about whether the animal we saw belonged to 'this species' or 'that species'. But that is tantamount to the original dispute only if we recognize the first picture *as* a picture of a cat and the second picture *as* a picture of a dog, in which case we are still implicitly using our recognitional capacities associated with the words 'cat' and 'dog' in thought, and so undermining the conditions of Chalmers' thought experiment, which depends on eliminating dependence on the terms at issue from thought as well as speech, otherwise any prototypically verbal dispute could survive the process, with each party tacitly relying on the ambiguous term in thought. In short, within Chalmers' setup, to count the original dispute as non-verbal, we must count 'cat' and 'dog' as expressing bedrock concepts, quite subverting his picture of bedrock concepts as forming a tiny, fundamental core. Obviously, 'cat' and 'dog' are not special in this respect; analogous arguments could be made about thousands of other natural or social kind terms.

Nor is the point restricted to kind terms. Imagine a dispute as to whether a paragraph of prose is 'subtle' or 'banal'. Only a philistine would think that such a dispute must be verbal. Yet, without the words 'subtle' or 'banal', the disputants may be unable to capture exactly what is at issue. Thus, 'subtle' and 'banal' come out expressing bedrock concepts too.

What has gone wrong with Chalmers' approach? He has neglected the association of much ordinary vocabulary with recognitional concepts. Eliminating a word typically involves eliminating the recognitional capacity that goes with it, for cats or dogs, or for the subtle or the banal. Yet much of our cognition depends on such recognitional capacities, deployed online to real-life cases or offline to hypothetical cases. Thus, the method of elimination does not leave us with the same cognitive powers, just honed down to fundamentals. Instead, it turns us into idiots, depriving us of most of the distinctions we need to think about the world as adults.

Why is the destructive effect of Chalmers' method of elimination not immediately obvious? The reason has to do with the phenomenon of *graceful degradation*, familiar in computer science. It is normally defined as something like the ability of a system to maintain limited functionality even when much of it has been destroyed or incapacitated. The point is to prevent catastrophic failure, or at least to postpone it for as long as possible. Graceful degradation is typical of systems that have evolved naturally or socially. A species of animals or plants which died as soon as they incurred injury or damage would not be evolutionarily robust. Natural languages too tend to degrade gracefully. They do not collapse when a single word is removed. Speakers find all sorts of workarounds. Thus, one can easily have the illusion that no 'essential' cognitive damage has been done. But, in reality, during Chalmers' process of iterated elimination, the language is being gradually eroded, bit by bit, in a sorites process. What is left at the end is just a stump. The cognitive power of a natural language consists in large part of a vast array of recognitional capacities associated with the terms of that language.[8]

References

Boghossian, Paul, and Timothy Williamson. 2020: *Debating the A Priori*. Oxford: Oxford
University Press.

Brown, Jessica. 1993: 'Natural kind terms and recognitional capacities', *Mind*, 107: 275–303.

Chalmers, David. 2011: 'Verbal disputes', *Philosophical Review*, 120: 515–566.

Davidson, Donald. 1977: 'The method of truth in metaphysics', *Midwest Studies in Philosophy*, 2: 244–254.

Dummett, Michael. 1973: *Frege: Philosophy of Language*. London: Duckworth.

Dunn, Jeffrey. 2014: 'Inferential evidence', *American Philosophical Quarterly*, 51: 203–213.

Grandy, Richard. 1973: 'Reference, meaning, and belief', *Journal of Philosophy*, 70: 439–452.

Huemer, Michael. 2024: *Ethical Intuitionism*. New York: Palgrave Macmillan.

McGrath, Sarah. 2004: 'Moral knowledge by perception', *Philosophical Perspectives*, 18: 209–228.

McGrath, Sarah. 2019: *Moral Knowledge*. Oxford: Oxford University Press.

Peacocke, Christopher. 2004: *The Realm of Reason*. Oxford: Clarendon Press.

Silk, Alex. 2017: 'Normative language in context', *Oxford Studies in Metaethics*, 12: 206–243.

Wedgwood, Ralph. 2006: 'How we know what we ought to be', *Proceedings of the Aristotelian Society*, 106: 63–86.

Williamson, Timothy. 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, Timothy. 2007: *The Philosophy of Philosophy*. Oxford: Wiley Blackwell.

8 Thanks to the audience at the second Balkan Analytic Forum for comments on an early version of this paper. The section on moral recognitional capacities is mainly based on chapter 2 of Williamson 2025.

Williamson, Timothy. 2020: *Suppose and Tell: The Semantics and Heuristics of Conditionals*. Oxford: Oxford University Press.

Williamson, Timothy. 2021: *The Philosophy of Philosophy*, enlarged ed. Oxford: Wiley Blackwell.

Williamson, Timothy. 2023: 'Knowledge-First Inferential Evidence', *The Monist*, 106: 441–445.

Williamson, Timothy. 2024a: *Overfitting and Heuristics in Philosophy*. New York: Oxford University Press.

Williamson, Timothy. 2024b: 'Where did it come from? Where will it go?', in Arturs Logins and Jacques-Henri Vollet (eds.), *Putting Knowledge to Work: New Directions for Knowledge-First Epistemology*, 21–70. Oxford: Oxford University Press.

Williamson, Timothy. 2025: *Good as Usual: Anti-Exceptionalist Essays on Values, Norms, and Action*. Oxford: Oxford University Press.

https://doi.org/10.18485/baf.2025.2.ch2

Slobodan Perović

THE COPERNICAN PRINCIPLE AND EVIDENCE IN MODERN COSMOLOGY

Abstract: Modern cosmology is not merely an empirical science but a domain deeply embedded in philosophical considerations. Various theoretical models of the universe that are underdetermined by evidence rest upon foundational principles, leading to epistemic and ontological debates about the status of these principles. The Copernican Principle is one example; it has played a crucial role, shaping the conceptual framework of cosmic structure and evolution. This preliminary critical examination of the Copernican Principle (Cosmological Principle) within the context of modern cosmological models explores its theoretical and evidential challenges and its indispensability. Although neither purely operational nor ontological, the principle has served as a key theoretical hinge to generate models and desired observational grounds within a field challenged by protracted underdetermination of models and theory by evidence.

Key words: cosmology; Copernican Principle; underdetermination; evidence

1. Cosmological Models

There is some controversy about when exactly cosmology was established as a modern scientific field (Kragh 1996). Some say it began with the 20th century development of the first cosmological models to apply relativistic field equations (successfully tested in the Solar System by Edington) to the level of the universe. Others point to the discovery of the cosmic microwave background (CMB) in 1965 (Penezias and Wilson 1965) confirming these models. What cannot be debated is that modern cosmology was built in the first two decades after WWII (Kragh 1996) on a foundation of competing models that sought to explain the origin, structure, and evolution of the universe during the so-called Great Controversy.

These models were not developed as theoretical constructs only, simply seeking to align with the available observations or to predict observed properties. They also purported to have profound philosophical implications and inherent principles that properly belonged to the domain of philosophical study, particularly their treatment of initial conditions, physical laws, and the nature of reality (Kragh 2013). This contention, plus the glaring underdetermination of cosmological models by data (Perović and Ćirković 2024, Ch 8; Ellis 2014), underscores the need for a careful philosophical analysis of the assumptions and principles underlying them.

1.1 Big Bang Models

The so-called Big Bang model, the cosmological framework that has emerged as the winner in the protracted cosmological "wars", first appeared in the mid-20th century in the work of scientists like Georges Lemaître, George Gamow, and Robert Dicke. De Sitter (1916) and Friedmann (1922) applied relativistic field equations at the cosmological scale prior to WWII. Lemaître's (1931) proposal of an initial "particle" that expands according to relativistic field equations as the origin of the universe laid the groundwork for the concept of an initial singularity from which the universe expands. Then, in the 1940s and 1950s, Dicke (1961) and Gamow (1949) developed a detailed model, with all the basic stages of the expansion of the universe. The model predicted the key observable properties, in particular, the microwave background radiation as a remnant of the primordial fireball of matter and radiation and its properties (Figure 1).

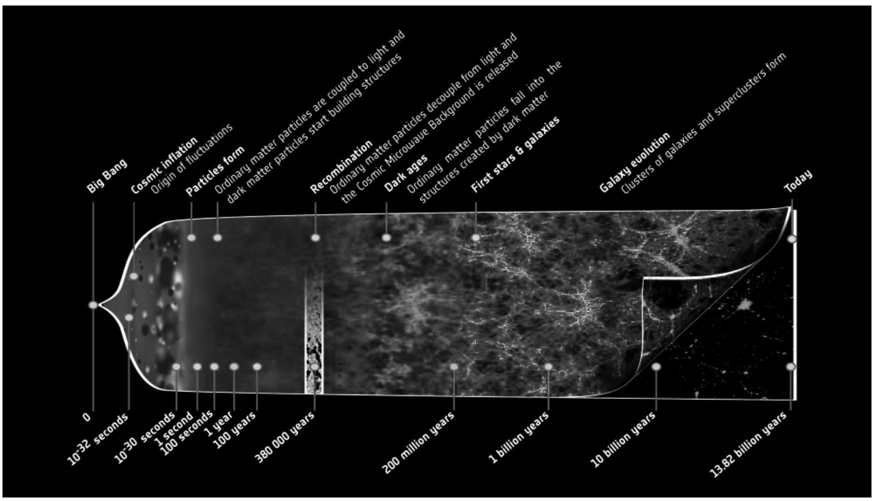


Figure 1: The basic stages of the evolution of the universe in the Big Bang model. (The diagram was originally posted as a Wiki-commons item.)

This predicted phenomenon – of seemingly strange 3K microwave radiation coming from all points on the sky, initially thought of by the discoverers as noise – was discovered in 1965, and after a protracted series of observations and arguments over its properties (temperature, isotropy, shape of the spectrum), it was finally confirmed as the remnant of the primordial fireball when detectors were placed on satellites in early 1990s (Perović and Ćirković 2024). We should note that this wasn't as easy as textbook accounts may suggest; various early Big Bang models, such as the Tepid Big Bang, Cold Big Bang, chaotic universe scenarios, and others also explored variations in initial conditions, each offering a different perspective on the universe's early state (Ibid).

It was also obvious from the start that the initial singularity raises deep epistemological questions as it does not imply the existence of space in any obvious way; rather, it seems to suggest a point where physical laws, particularly General Relativity, may break down. The epistemic status of this singularity triggers fundamental questions about whether it represents a genuine feature of reality or a limitation of current physical theories. The philosophical problem of why the universe began with particular initial conditions remains an open question, influencing contemporary debates on fine-tuning and multiverse theories (Ellis 2014).

1.2 Steady State Models

In contrast to the Big Bang model(s), the Steady State models of the universe, championed by figures like Fred Hoyle, Hermann Bondi, Thomas Gold, and others, posited an infinite and unchanging universe. These models, prominent from the 1930s to 1960s and remaining alive until the 1980s, argued for a constant density of matter over time, achieved through the continuous creation of matter, and the steady expansion of an infinite universe. They dealt with the continuous origination of matter in various ways. Most prominently, Fred Hoyle (1948) introduced a universal scalar creation field (C-field) as the mechanism providing the non-collapsing steadily expanding dynamics of the infinite universe that otherwise would eventually decelerate and collapse due to gravitational pull – as the Big Bang models initially predicted. The Steady State models' explanation of the spectral redshift (shift to the red region of the spectrum) of the receding galaxies was pretty much as convincing as that of the Big Bang models, so the multi-decade controversy was based, in part, on the interpretation of this key piece of evidence – at least until 1965 and the discovery of the CMB.

The Steady State framework for models explicitly relied on the assumption that physical laws are immutable, thereby avoiding the need to explain contingent initial conditions (as there were none) (Bondi and Gold 1948, Gregory 2005, Perović and Ćirković 2024, Ch 18). This assumption reflected a deeper epistemic commitment to the idea that the universe should not exhibit special initial conditions, and any principle underlying cosmological models should stick to epistemologically sound views.

1.3 Alternative Models

Unlike the Big Bang and Steady State models, other cosmological frameworks, such as Dirac's cosmologies (Dirac 1974; Perović and Ćirković 2024, Ch15), suggested physical constants might vary across time and space. Roger Penrose (1979) proposed the arrow of time arises from specific initial conditions rather than fundamental laws. These alternatives highlight the philosophical tension between necessity in physical laws and contingency in cosmological explanations, raising questions about the nature of physical laws themselves and their dependence on boundary conditions. But they do so in a rather different direction than the Steady State models.

2. The Copernican (Cosmological) Principle

In the foundational discussions of the models mentioned above, one principle stands out. The Copernican Principle, a cornerstone of modern cosmology, asserts that the universe is homogeneous and isotropic – in other words, it appears the same in all directions and all locations (Weinberg 1972). Thus, the density of radiation and matter cannot vary greatly, nor will it matter at which point one measures these values.

The principle has both strong and weak formulations, each with distinct philosophical and scientific implications relevant to the cosmological models at stake. The strong version, commonly known as Perfect Copernican (or Cosmological) Principle posits the universe is homogeneous and isotropic *at all times* and is thus characterized by constant density. This principle underpins all Steady State models, given their attempts to avoid the apparent contingency of the initial conditions of Big Bang cosmology. By assuming no cosmological feature escapes the laws of physics, the strong version of the Copernican Principle reflects an epistemological motivation to maintain regularity and predictability in the universe. (Bondi and Gold 1948, Hoyle 1955, Bondi 1960) Bondi and Gold provided a deduction into the cosmological models from the Perfect Cosmological Principle (i.e., the strong version of the principle).

The weaker version however, nowadays often simply equated with the label of 'Copernican Principle' without qualifiers due to the prevalence of the Big Bang model, allows for changes in density over time due to cosmic expansion. This version is compatible with Big Bang models describing a dynamic universe with evolving properties. Thus, the universe is homogeneous and isotropic *at any given tim*e, but not at all times, as it expands.

The Copernican Principle has been explicitly or implicitly challenged or rendered irrelevant by the alternative models, such as the models proposed by Dirac and Penrose suggesting homogeneity and isotropy may not hold universally. In Dirac's and Penrose's model, the Copernican Principle is treated as tenuous, if not dispensable, given that the physical laws themselves are either changeable or the result of arbitrary initial conditions.

The immediate philosophical challenge here is to identify what kind of principle the Copernican Principle actually is. Is it an empirical generalization to be tested, or some sort of necessary constraint on cosmological models? (Beisbart 2009) It is certainly treated as the former by astrophysicists (Camarena, Marra and Clarckson 2022), but its ontological status remains contested (Ellis 2014, Beisbart 2009, Besibart and Jung 2006). Does it have that sort of status at all? Is it a mere methodological assumption, or does it reflect an objective feature of reality?

Instead of going into discussion head-on I think we can set the stage for addressing and answering these questions by understanding the relationship between *the evidence and the principle*, both in practice and normatively. It is part of a more general question of how cosmologists treat principles such as the Copernican Principle in light of novel evidence, and what epistemic standards of the community ought to govern such a relationship. Examining this relationship will offer a unique perspective that may be more informative than simply looking at the relationship between the principles and the models.

3. Cosmological Models and Evidence

From the beginning of the 20th century and the development of relativistic cosmological models, the development of cosmological models has been shaped by ongoing efforts to reconcile theoretical predictions with observational evidence. The discovery of the CMB in 1965 provided strong support for the Big Bang theory, leading to the decline of Steady State models. However, the refinement of these models over four decades highlights the protracted scientific and epistemic battle between these competing frameworks and more than just transitory nature of underdetermination in the field (Perović and Ćirković 2024).

Cosmological evidence is primarily observational, not experimental, relying on signals from the deep past, such as the CMB. Unlike, for instance, an experimental field of particle physics, cosmology's evidential basis is observational and strongly model-dependent. This raises a question about the extent and space for interpretation of data and the role of theoretical principles in shaping scientific understanding across fields. The reliance on indirect evidence from the deep past – in fact as deep as naturally possible as far as we know – makes cosmology particularly susceptible to underdetermination of theories and models by evidence and to theory-ladenness (Perović and Ćirković, Ch. 8; Butterfield 2014; Ellis 2014). Even if underdetermination is transitory, it is protracted in cosmology in comparison to the fields where key parameters of interest can be controlled and varied in experimental conditions. In such a domain of available evidence, plausible and precise alternative models have been built, improved upon, and adjusted for decades. This applies to each key piece of evidence, be it redshift, CMB properties, or deceleration parameters. In a sense, the protracted proliferation and testing of various alternative models is a sign of a fruitful field if confined to observational, rather than predominantly experimental evidence.

4. The Copernican Principle and Evidence

4.1 The Copernican Principle as a theoretical hinge

The Copernican Principle's use as a guiding principle across cosmological models raises a philosophical dilemma. Is it a purely operational tool, and if so, how exactly does it "operate" within models? Or does it have unavoidable ontological features and implications indirectly connected to the models' parameters? Does any ontological feature go over and above the operational roles? Is this relevant to a particular model, and if so, how?

As I hinted in Section 3, instead of addressing the questions head-on here, I am taking a roundabout route. The goal is to determine what sort of role the Copernican Principle plays in the context of available or projected (by the model at stake) observational evidence. We can understand what the principle is if we understand how it functions in practice and how it ought to function given the nature of the scientific pursuit of cosmology as a scientific field.

First, undeniably, in cosmological work itself, the debates on the Copernican Principle have been couched as ontological (Perović and Ćirković 2024; Kragh 1996), and this, in turn, has defined the models' explanatory features. The debates and dilemmas stemming from them have been a staple of modern cosmological frameworks and debates among their proponents. This suggests the role of principles in cosmology extends beyond their operational utility, and that deductions concerning them have an impact on how models are created. But should we take the pronouncements and characterizations we find in the actual debates for granted? Are they perhaps only rhetorical devices?

Answering this question is a hefty task, with some sound philosophical analysis preceding the question concerning particular models (Beisbart 2009, Beisbart and Jung 2006), but if we stop using the usual philosophical labels for the moment, it seems clear that the Copernican Principle, as well as other similar principles, serves as a theoretical hinge for generating novel interpretations and models. It is initially formulated in a strong or weak fashion and then admitted into models as plausible by being connected with an operational parameter of some sort. Let me briefly state the case for this view.

The creators of both classes of early Steady State models argued vehemently for the strong version of the Copernican Principle (Bondi and Gold 1948; Gregory 2006; Kragh 2016; Kragh 2011). They not only thought it was plausible but they also either explicitly or implicitly pointed out what they argued were bad or even unacceptable implications of the alternatives. In this view, any principle that leads to the understanding of the laws of physics as immutable undercuts any attempt to understand the physical nature of the universe. The laws may be mutable, and if they are, we ought to give up attempts to understand the universe by means of physics. Thus, the weak version was not only judged as epistemically admissible and the preferred foundation of cosmological models, but was mostly regarded as the only viable option throughout the multi-decade development of various versions of Steady State models. Meanwhile, the key operational parameters of these models related to steady expansion and infinity, i.e., the properties of matter and radiation, were *aligned* with the strong version of the principle. Moreover, new parameters, such as the C-field, were introduced, and the observational parameters were interpreted along the lines of the models' key operational parameters aligned with the principle.

In fact, the proponents of the Steady State models (Bondi and Gold 1948) put weight on the restrictiveness of the Copernican Principle for ontological and epistemological reasons: to close what they regarded as the nonviable ontological possibility the weaker version opens up – the possibility of mutable physical laws – in light of the essential unknowability of the universe, that they thought contradicts our experience. The ontological concern and the formulation of the principle were thus pretty directly transferred through the red line of epistemic admissibility into the models' parameters. It's quite possible to speculate that the model could have been created without the deduction from the principle, but the fact is that the models were perfectly aligned with the Perfect Copernican Principle all along, and explicitly so. The speculation of that sort would be rather explanatorily vacuous.

The proponents of the Big Bang models followed somewhat different reasoning (Weinberg 1972). The weak version of the Copernican Principle was treated as an admissible foundation of the model, but its acceptance was more a consequence of weighing its plausibility against other features of cosmological models. The alternative, the strong version, was regarded as plausible but not inescapable. The operational parameters related to the expansion and density of matter and radiation over time were consistent with the weak principle but emphasis was put on the plausibility of the model's features and its increasing alignment with observations, rather than on the plausibility of the weak Copernican Principle *per se*. Yet this opened a can of worms: the status of initial singularity potentially fell out of the explanatory domain of the General Theory of Relativity, as I mentioned previously.

Moreover, after the 1980s, it turned out, precisely due to the shocking (to the Big Bang Proponents) observational result against which the principle was weighed, that the universe is accelerating, not decelerating as the proponents of the Big Bang models expected (simply due to the eventual gravitational pull of matter into itself – an implication that even Newton anticipated). The most convincing explanation of this result so far has resorted to the scalar inflationary field (Guth 1981, Linde 1990), a move not only reminiscent of, but inspired by, the scalar C-field introduction into the Steady State models (Gregory 2005, 324). The universe is not steadily expanding, as the proponents of the Steady State argued, but "something" makes it accelerate. And this "something" was conceptualized the way the C-field was. Thus, the observations eventually forced the Big Bang proponents to opt for the same kind of parameter (although of a different value), although most had previously argued this was an epistemically unsound if not entirely inadmissible move. This seems to suggest a rather hasty and more naïve epistemological attitude among Big Bang proponents than Steady State proponents, despite their triumph.

To make things even more complicated, the Dirac and Penrose cosmologies basically start with the open epistemic admissibility criterion as central and essentially treat the Copernican Principle as tenuous, to the point of making it only locally relevant. Dirac's model implies the mutability of physical laws is a plausible and preferable hypothesis that should be tested (Kragh 1982), while Penrose's model disposes of the notion that we are detecting universal physical laws at all, on cosmological grounds. Our universe is just a universe in the chain reaction of the creation of universes, the local boundary conditions of which determine what appear to us to be universal laws. Penrose suggested ways of testing his model, not just how existing observations aligned with it.

Thus, rather than characterizing it as purely operational, or ontological only in name, if we are true to the actual role the Copernican Principle has played in cosmological arguments, then we must accept that at least minimally, it has served as a theoretical hinge to justify or even generate operational parameters in novel models and interpretations and to generate novel sorts of observations or adjust models to novel observations (Figure 2). It has done so in different ways in terms of weighing various steps across models.

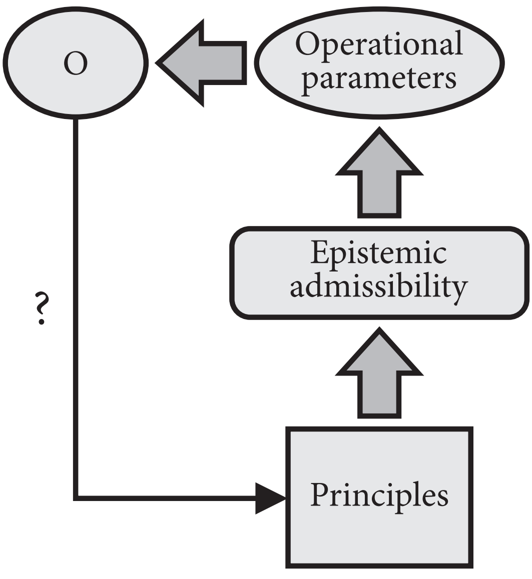


Figure 2: Various abstract principles are deemed as admissible or inadmissible into the cosmological models, on their own terms, then co-aligned with the operational parameters; finally, the latter are tested observationally. Whether the weight of the observational evidence suffices to give up the principles is the key question we consider.

4.2 Epistemological responsibility in cosmology, and theoretical hinges

This brings us to the second issue. When examining the status of a principle by looking at how it relates to observations, albeit indirectly via the operational parameters, in terms of epistemic responsibility when creating a model, at what point should a principle be abandoned in light of conflicting evidence? What strength of evidence is required to challenge a seemingly foundational principle, and thus the model it founds, via tested operational parameters?

The reasons for accepting or abandoning a principle in light of particular evidence offered by individual cosmologists can be analyzed on a case-by-case basis. But a more important general epistemological issue concerns the epistemological standards that ought to govern the community's treatment of models based on principles such as the Copernican Principle. When should they be accepted and when abandoned?

The return of ideas, such as the transition from the C-field in Steady State models to inflationary fields in Big Bang cosmology, is not an exception, although this particular case stands out. Generally speaking, in cosmology, models are underdetermined by evidence for protracted periods of time; thus, "fringe" alternatives that support or abandon established principles occupy a unique epistemic status. They have a special status compared to heavily experimental fields, where they are often justifiably treated as refuted and then abandoned (Perović 2021). The "tolerance" of an outlier that either holds on to a version of the Copernican Principle or abandons it wholesale should be a priority. Prolonged underdetermination requires diverse theoretical hinges and their application.

The interplay between principles and evidence in cosmology is obviously quite dynamic, but this evolution reflects broader questions in the philosophy of science about the persistence and transformation of theoretical commitments over time. Now, even in High Energy Physics, a heavy-handed experimental field, the situation is mixed, so to speak. During the first phase of experiments with colliding particles, the key parameters, such as energy, momentum, various charges, or angle of collision, can be manipulated at will. Yet after the collision, when the collided beams spray vast amounts of debris, sifting through the debris and analyzing it is very much in the domain of observational activity similar to astrophysicists scanning vast numbers of stars, clusters, and other celestial objects and inferring conclusions from their observations. The epistemological challenges in High Energy Physics may not be at the scale of cosmology due to underdetermination and observational uncertainties, but they are there nonetheless and worth thinking about for philosophical and pragmatic reasons.

5. Conclusion

The philosophical features of modern cosmology, as exemplified by the case of the Copernican Principle, reveal an intricate relationship between scientific principles, evidence, and philosophical inquiry. Cosmological models, from the Big Bang to Steady State theories, have reflected competing visions of the universe's nature and origins, while the underdetermination of cosmological theories by evidence and the evolving nature of observational data underscore the provisional status of even the most foundational principles. In such a context, the Copernican Principle, in its strong and weak forms, serves as a basis for models' construction and evaluation.

References

Beisbart, C. (2009). Can we justifiably assume the cosmological principle in order to break model underdetermination in cosmology?. *Journal for general philosophy of science*, *40*, 175–205.

Beisbart, C., & Jung, T. (2006). Privileged, Typical, or not even that?–Our Place in the World According to the Copernican and the Cosmological Principles. *Journal for General Philosophy of Science*, *37*, 225–256.

Bondi, H. (1960). *Cosmology*. Cambridge: Cambridge University Press.

Bondi, H., & Gold, T. (1948). The steady-state theory of the expanding universe. Monthly Notices of the Royal Astronomical Society, 108(3), 252–270.

Butterfield, J. (2014). On under-determination in cosmology. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 46, 57–69.

Camarena, D., Marra, V., Sakr, Z., & Clarkson, C. (2022). The Copernican principle in light of the latest cosmological data. *Monthly Notices of the Royal Astronomical Society*, 509(1), 1291–1302.

De Sitter, W. (1916). "Einstein's Theory of Gravitation and Its Astronomical Consequences." *Monthly Notices of the Royal Astronomical Society* 76: 699–728.

Dicke, R. H. (1961). Dirac's Cosmology and Mach's Principle. *Nature* 192, 440–441.

Dirac, P. A. M. (1974). Cosmological models and the large numbers hypothesis.

Proceedings of the Royal Society of London A 338, 439–446.

Ellis, G. F. R. 2014. "On the Philosophy of Cosmology." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 46: 5–23.

Friedmann, A. 1922. "Über die Krümmung des Raumes." Zeitschrift für Physik 10, no. 1: 377–86.

Gamow, G. (1949). On relativistic cosmogony. *Reviews of Modern Physics* 21, 367–373.

Gregory, J. (2005). *Fred Hoyle's Universe*. Oxford: Oxford University Press.

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23(2), 347.

Hoyle, F. (1948). A new model for the expanding universe. *Monthly Notices of the Royal*

Astronomical Society (MNRAS) 108, 372–382.

Hoyle, F. (1955). *Frontiers of Astronomy*. New York: Perennial Library.

Kragh, H. (1982). Cosmo-physics in the thirties: Towards a history of Dirac cosmology. *Historical Studies in the Physical Sciences*, 13(1), 69–108.

Kragh, H. (1996). *Cosmology and Controversy*. Princeton: Princeton University Press.

Kragh, H. (2011). *Higher Speculations: Grand Theories and Failed Revolutions in Physics*

and Cosmology. Oxford: Oxford University Press.

Lemaître, G. (1931). The beginning of the world from the point of view of quantum theory.

Nature 127(3210), 706.

Linde, A. D. (1990). *Inflation and Quantum Cosmology*. Boston: Academic Press.

Penrose, R. (1979). Singularities and time-asymmetry. In S. W. Hawking, & W. Israel

(eds.), *General Relativity: An Einstein Centenary* (pp. 581–638). Cambridge: Cambridge

University Press.

Penzias, A. A., & Wilson, R. W. (1965). A measurement of excess antenna temperature at

4080 Mc/s. *Astrophysical Journal* 142, 419–421.

Perović, S. (2021). Observation, Experiment, and Scientific Practice. *International Studies in the Philosophy of Science*, 34(1), 1–20.

Perović, S., & Ćirković, M. M. (2024). *The Cosmic Microwave Background*: Historical and Philosophical Lessons. Cambridge University Press.

Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the*

General Theory of Relativity. New York: Wiley.

2. DISPOSITIONS AND PERSONALITY TRAITS

https://doi.org/10.18485/baf.2025.2.ch3

Aleksandra M. Pavlović

DISPOSITIONS AND BRAIN HEALTH: THE ROLE OF PREMORBID PERSONALITY TRAITS IN STROKE RISK ASSESSMENT

Abstract: Mental health and brain health have frequently been regarded as distinct dominions. Nevertheless, the complex integration of mental and brain health encompasses numerous shared pathways that affect both psychiatric and neurological disorders. Poor physical health significantly impacts neuropsychiatric conditions and may be a cause of some: for example, stroke can be a consequence of heart disease, brain and spinal cord can be affected by vitamin B12 deficiency due to gastritis, etc. On the other side, psychiatric conditions have been recognized as risk factors for somatic conditions, such as asthma or gastrointestinal and cardiovascular disorders.

Stroke is the second leading cause of death and the primary cause of long-term disability globally. Personality traits emerged as an important factor in both stroke recovery and adherence to medications and lifestyle modification. Personality and behavioral changes have been widely studied after traumatic brain injury and stroke and have been associated with lesions of the frontal and temporal lobes. After a stroke, patients often present with maladaptive personality traits associated with negative affect such as anxiety, emotional lability and rigid perfectionism, and interpersonal problems, which correlate positively with avoidant, dependent, and obsessive-compulsive personality disorders.

Although stroke risk factors have been well-studied for decades, they cannot be detected in all cases, particularly in younger patients in whom stroke incidence is steadily increasing worldwide. Therefore, researchers are now exploring novel modifiable risk factors, such as inflammatory disorders, infections, air pollution, silent cardiac disorders, and heart rate and blood pressure variability. Pre-stroke personality traits can be added to this list according to recent research. Data from six longitudinal cohorts including over 58,000 individuals monitored long-term showed that elevated levels of neuroticism were associated with an increased risk of stroke, while higher levels of conscientiousness were found to be protective against stroke. The combination of higher neuroticism and lower conscientiousness was also associated with a steeper cognitive decline and a higher risk of Alzheimer's disease and vascular dementia. Although Alzheimer's disease and vas-

cular dementia have distinct etiologies, they share common features, including vascular risk factors and brain vascular pathology, which contribute to cognitive decline in both conditions. Emerging evidence highlights an intriguing association between personality traits (e.g., higher neuroticism and lower conscientiousness) and both Alzheimer's disease and vascular dementia, potentially mediated through shared vascular risk factors and neuropathological changes, despite their distinct etiological origins. These personality traits have also been related to vascular risk factors, higher inflammation and mortality.

Recent evidence suggests that the etiology of psychiatric and neurodegenerative disorders may partially originate in early neurodevelopment, explained by a complex and still largely unknown interplay of genetic and epigenetic factors. The relationship between the health of the mind, brain, and body seems to be multifaceted and reciprocal, and yet intuitive. However, the is exact mechanisms unerlying these connections are not yet clear, particularly in the context of disease. There is a need for triangulation of psychological, psychiatric, and neurological aspects of overall body health in large, well-designed studies.

Introduction

Mental health and brain health have frequently been considered distinct domains in medicine and psychology. Still, the complex integration of mental health and brain functioning encompasses many shared pathways that can be identified in numerous psychiatric and neurological disorders (Ibanez & Zimmer, 2023). Poor physical health can lead to neurological diseases and can significantly impact psychiatric conditions. Likewise, psychiatric disorders can be associated with somatic symptoms and diseases in both adults and children (Agnafors et al., 2019; Ibanez & Zimmer, 2023). There are indications that the origins of psychiatric and neurodegenerative disorders, which have typically been categorized separately, may partly stem from early neurodevelopment (Thibaut, 2018). A better understanding of the intricate interactions between these two domains must be based on the exploration of relationships between genetics and epigenetics, the effects of social determinants and environmental pollutants, and the cumulative impact of our lifestyle choices (Craske et al., 2023).

Stroke is the second leading cause of death and the primary cause of long-term disability globally, but its epidemiology varies significantly by region, age, sex, and socioeconomic status (Cordonnier et al., 2017; GBD 2019 Stroke Collaborators). Stroke is not a single disease but a heterogeneous condition in terms of presentation and etiology, and it can be caused by a wide range of risk factors, pathological processes and mechanisms (Murphy & Werring, 2020). It is largely classified into two types:

ischemic stroke, caused by an obstruction in blood flow to the brain, and hemorrhagic stroke, resulting from bleeding in the brain. Globally, ischemic strokes account for approximately 85% of all cases. The incidence of stroke is increasing in low– and middle-income countries, partly due to rising rates of hypertension, diabetes, and lifestyle factors such as poor diet and physical inactivity, which are well-known risk factors for all subtypes of stroke (GBD 2019 Stroke Collaborators). Although stroke incidence and mortality have declined in many high-income countries due to improved medical care and prevention strategies, the incidence is steadily rising in younger patients globally (Ma et al., 2024). Worldwide, stroke remains a critical public health issue, necessitating continued efforts in risk factor management and healthcare access improvements.

Personality traits are relatively enduring emotional, cognitive, and behavioral patterns that characterize people over time, and influence an individual's interactions and adaptability to their environment (Costa et al., 2019). Personality traits are relatively consistent over time, particularly during midlife, which is the most relevant life period for stroke and dementia prevention (Larsen & Buss 2017; Lucas & Donnellan, 2011). The Big Five personality traits, also known as the five-factor model, are widely utilized in clinical and health psychology to describe personality characteristics, the core dimensions being "extraversion," "agreeableness," "conscientiousness," "neuroticism," and "openness to experience" (McCrae & John, 1992). Extraversion pertains to the extent of an individual's interpersonal engagement and tendencies towards sociability and talkativeness. Agreeableness reflects the propensity for harmonious and comfortable interpersonal relationships. Conscientiousness involves a commitment to societal norms and principles. Neuroticism indicates levels of emotional instability, while openness encompasses a preference for diversity, change, and intellectual exploration.

In this paper, I will explore the role of personality in stroke recovery and prevention, as well as the evidence linking certain personality traits to an increased risk of stroke. I will also review the current state of research on personality changes following brain injury, with a particular focus on stroke.

The search for new stroke risk factors

Risk factors for stroke have been extensively studied over the past few decades, leading to the identification of two main and well-established categories: modifiable and nonmodifiable risk factors. Age, sex, and race/ethnicity are nonmodifiable risk factors for both ischemic and hemorrhagic stroke, while hypertension, smoking, diet, and physical inactivity are among modifiable risk factors (Boehme et al., 2017). Since these diseases and conditions explain the majority but not all stroke events, there is an ongoing effort to expand our understanding of stroke mechanisms and identify novel risk factors (Pavlovic, 2021). More recently documented risk factors and triggers of stroke include inflammatory disorders, infection, air pollution, heart-rate variability, and cardiac atrial disorders unrelated to atrial fibrillation (Boehme et al., 2017; Pavlovic, 2021). Identifying modifiable risk factors offers an opportunity to improve both primary prevention (reducing the risk of an incident stroke) and secondary stroke prevention (reducing the risk of recurrent strokes). For example, weight reduction, low-risk diet, regular physical activity, smoking cessation, and low-to-moderate alcohol consumption may reduce stroke risk by up to 50% or more, but level-one evidence is still lacking for several of these interventions (Niewada & Michel, 2016).

Stroke consequences

In many survivors, stroke may result in life-changing consequences, as two-thirds show motor or cognitive disability leading to limitations in activities of daily living (Sennfält et al., 2019; Gil-Salcedo et al., 2022). Psychological and behavioral changes after stroke have negative effects on functional recovery, increasing the rate of mortality and disability of survivors (Zhou et al., 2023). In the acute stroke setting, patients can become disinhibited, agitated, or aggressive, and anxiety and depressive symptoms are quite frequent (Zhang et al., 2020). Long-term cognitive and neuropsychiatric effects of stroke, including behavioral impairments such as poor motivation, apathy, and a tendency to be self-centered and less aware of the needs of others, are commonly reported (Zhang et al., 2020). There is also a higher risk of dementia, which is attributed to lesion location but also preexisting brain changes and pre-stroke cognitive decline (Pendlebury & Rothwell, 2009; Levine et al., 2015; Mijajlović et al., 2017; Kuźma et al., 2018; Verdelho et al., 2021).

Personality changes and acquired brain lesions

Undoubtedly, personality changes after acquired brain injury of any etiology can profoundly impact quality of life by affecting social, familial, and professional functioning (Halalmeh et al., 2024). However, there is also a controversial concept of posttraumatic growth, which refers to enduring positive psychological change experienced as a result of adversity, trauma, or highly challenging life circumstances (Jayawickreme et al., 2021). Notably, an individual's ability to adjust and cope with changes associated with brain injury is linked to premorbid personality factors (Podell et al., 2010). Premorbid conditions may play a part in personality changes, as several studies suggest that social and emotional deficits or personality variations may be explained by preexisting susceptibilities and characteristics. For example, in a study by Greve and coworkers, a large majority of the individuals with impulsive aggression traits had premorbid personality traits of aggression and behavior that would be categorized as impulsive, suggesting that brain injury did not cause the impulse aggression but rather exacerbated premorbid characteristics in the now disinhibited patients due to frontal lobe damage (Greve et al., 2001; Rieger, 2015). In addition, some behaviors and personality traits may be associated with injury leading to brain damage, such as in the case of a history of substance abuse, which is a risk factor for accidents resulting in injury (Rieger, 2015).

Personality changes due to stroke

Personality change is an important psychiatric complication following a stroke, and is linked to severe affective dysregulation and behavioral alterations (Lo Buono et al., 2022). Studies on personality change after stroke are generally lacking. Personality and behavioral changes after traumatic brain injuries have been widely studied, and are associated with the damage to the frontal and temporal lobes, which appear to be the most susceptible to damage in the setting of blunt force head trauma (Stone et al., 2004; Rieger, 2015). More than twenty years ago, Bogousslavsky emphasized the importance of negative symptoms in stroke patients, including sadness, disinhibition, lack of adaptation, environmental withdrawal, crying, passivity, and aggressiveness (Bogousslavsky, 2003). After a stroke, patients often present with maladaptive personality traits associated with negative affect, such as anxiety, emotional lability and rigid perfectionism, and interpersonal problems, which are correlated positively with cluster C personality disorders, including the avoidant, dependent, and obsessive-compulsive personality disorders (Lo Buono et al., 2022). Personality changes after a stroke are likely a consequence of brain damage that disrupts neural circuits regulating emotions and behavior, particularly in the frontal lobes (responsible for impulsivity and decision-making) and temporal lobes (involved in memory and communication) (Kim, 2016; Mijajlović et al., 2017). Lo Buono and coworkers describe this negative affectivity after stroke as an experience of negative emotions and dysfunctional coping (Lo Buono et al., 2022). Maladaptive traits, such as neuroticism, may impede recovery from brain lesions, while protective features such as extraversion and conscientiousness can facilitate the rehabilitative process (Waszczuk et al., 2018). The most frequent behavioral and personality changes after stroke are summarized in Table 1. However, even in patients without clinically evident stroke, the presence of chronic vascular brain changes is associated with development of the depression and apathy (Pavlovic et al., 2016).

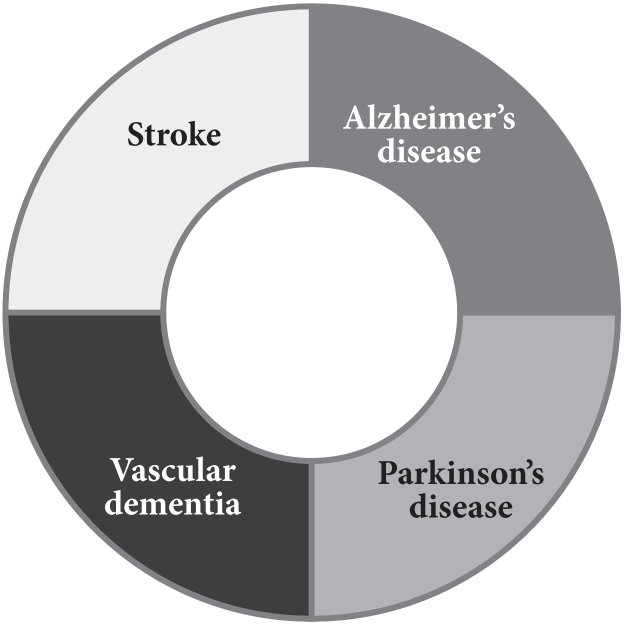
Table 1. Behavioral and personality changes after stroke

Phenomenon	Description
Apathy	lack of motivation or loss of interest in life
Depression	depressed mood, loss of interest and energy
Anxiety	extreme and unreasonable fear and difficulty managing daily tasks
Confusion/Delirium	altered state of consciousness, characterized by episodes of confusion
Pseudobulbar affect	emotional incontinence, unpredictable episodes of crying or laughing, often in socially inappropriate situations
Impulsiveness	inability to think ahead or understand consequences
Irritability	frustration or loss of temper
Fatigue	extreme tiredness developed during mental activities

Personality and stroke risk

Stephan and colleagues (2023) employed a multi-cohort design to investigate relationships between the Five-Factor Model personality traits and the occurrence of stroke. This study used data from six large-scale longitudinal cohorts, encompassing 58,105 individuals aged 16 to 104. The incidence of stroke was monitored over a follow-up period ranging from 7 to 20 years. The meta-analysis revealed that elevated levels of neuroticism were associated with an increased risk of stroke (hazard ratio [HR] = 1.15; P <0.001), while higher levels of conscientiousness were found to be protective against stroke (HR = 0.89; P <0.001) (Stephan et al., 2023). Further analysis showed that these associations were partially mediated by body mass index, diabetes, blood pressure, physical inactivity, and smoking (Stephan et al., 2023). Openness has been associated with an increased risk of stroke (Weston et al., 2015), which is interesting given the reported positive association between openness and longevity (Turiano et al., 2012). In a study by Weston and collaborators, a one unit increase in openness decreased the odds of a stroke diagnosis by 31%, heart conditions by 17%, and high blood pressure by 29% (Weston et al., 2015). In this large dataset, Stephan and coauthors did not find significant relationship between traits of extraversion, openness, and agreeableness and stroke incidence compared to earlier studies with smaller number of participants (Stephan et al., 2023).

The pattern of higher neuroticism increasing stroke risk and higher conscientiousness protecting against stroke was also reported for cardiovascular and some other neurological conditions (Graph 1) (Wright et al., 2022; Stephan et al., 2023). This combination of higher neuroticism and lower conscientiousness was associated with a steeper cognitive decline (Sutin et al., 2023), a higher risk of Alzheimer's disease (Aschwanden et al., 2021), and an increased risk of vascular dementia (Terracciano et al., 2021a). An association between higher neuroticism and a higher risk of Parkinson's disease has also been reported (Terracciano et al., 2021b). Lower levels of agreeableness predicted poorer physical and subjective health (Turiano et al., 2012) and premature mortality (Jackson et al., 2015).



Graph 1. Neurological diseases most consistently associated with higher neuroticism and lower conscientiousness

Potential mechanisms linking personality traits to stroke risk

Recent findings by Terracciano and co-workers on data from the UK Biobank comprising 40,602 middle-aged and older adults showed that participants who scored higher on neuroticism had more white matter hyperintensities ($\beta = 0.024$; $p < 0.001$), which was consistent across periventricular and deep brain regions (Terracciano et al., 2023). White matter hyperintensities in these age groups are typically due to chronic vascular changes affecting small brain vessels, and may also be a marker of alterations of white matter integrity. Interestingly, the association was reduced by about 40% when accounting for vascular risk factors (smoking, obesity, diabetes, high blood pressure, heart attack, angina, and stroke) (Terracciano et al., 2023). This association was not moderated by age, sex, college education, deprivation index, or APOE e4 genotype, and remained unchanged in sensitivity analyses that excluded individuals with dementia and those younger than 65 (Terracciano et al., 2023). In addition, the authors found that white matter hyperintensities partly mediated the association between neuroticism and cognitive function (Terracciano et al., 2023).

In terms of potential mechanisms underlying these associations, personality traits have been linked to vascular risk factors (Weston et al., 2020), inflammatory biomarkers (Wright et al., 2022), and stress responses (Leger et al., 2021). Higher neuroticism and lower conscientiousness, in particular, are associated with increased cardiovascular risk (Čukić & Bates, 2015; Thomas et al., 2022) and stroke risk factors, including a higher risk of heart disease (Weston et al., 2020), elevated body mass index (Sutin & Terracciano, 2016), and an increased risk of diabetes and hypertension (Weston et al., 2020). Additionally, these traits are linked to behaviors such as smoking (Hakulinen et al., 2015) and physical inactivity (Sutin et al., 2016). Wright and colleagues found that higher neuroticism and lower conscientiousness are associated with greater systemic inflammation (Wright et al., 2022) which is also a known risk factor for stroke. High levels of conscientiousness have been connected to lower levels of IL-6 and CRP (Graham et al., 2018; Turiano et al., 2013), while neuroticism has been linked to higher levels of these inflammatory markers, though the associated effect sizes are generally small (Graham et al., 2018; Sutin et al., 2010). Leger and coauthors suggested that negative affect reactivity to everyday stress may also partially explain the association between higher neuroticism and lower conscientiousness (Leger et al., 2021). Interestingly, reported contradictory pattern of a relationship between cortisol and personality variables, especially neuroticism, can partially be explained by sex differences as for men neuroticism was positively correlated with cortisol level, while for women it was negatively correlated (DeSoto & Salinas, 2015). Different personality traits have been associated with individual biological reactivity to stressful events by many biomarkers of stress-sensitive biological systems, not only including cortisol levels, nicely summarized in paper by Soliemanifar and coworkers (2018).

Most vascular risk factors – such as hypertension diabetes, and high cholesterol – are chronic diseases and their management requires long-term pharmacological therapy. Medication adherence is typically defined as the degree to which patients follow their healthcare provider's instructions regarding their medication, and is influenced by various sociodemographic, psychological, and clinical factors (Osterberg & Blaschke, 2005; Ko et al., 2020). Medication adherence also involves patients' understanding of medications and active participation in treatment decisions (Adachi et al., 2022). Patients with certain personality traits may find adherence to medication and recognition of a disease challenging (Bailey, 2016). Previous studies have suggested that adherence can be related to personality traits in several contexts, including chronic conditions like rheumatoid arthritis, HIV, and multiple sclerosis, and medical treatments like hemodialysis and cholesterol-lowering therapy (Wiebe & Christensen, 1997; Stilley et al., 2004; O'Cleirigh et al., 2007; Bruce et al, 2010; Leon et al, 2016; Santos et al., 2019; Ko et al., 2020). In particular, in patients with cardiovascular conditions, a higher conscientiousness score was significantly associated with three of four major aspects of medication adherence: a high medication compliance score (taking medications as prescribed), a high score on collaboration with healthcare providers, and a high score on willingness to access and use information about medications. Conscientiousness, extraversion, and agreeableness were the most commonly reported personality traits associated with good adherence, while neuroticism was associated with medication non-adherence (Wheeler et al., 2012; Molloy et al., 2014; Kohli, 2017). In addition, there is also evidence of an association between low conscientiousness and a high dropout rate of cardiac rehabilitation in patients with cardiovascular diseases (Adachi et al., 2021).

These findings will likely be expanded soon to include a broader range of neurological conditions. In a recent study, Zhang and coauthors analyzed the relationship between personality traits (such as warmth, diligence, nervousness, sociability, and curiosity), and brain disorders in 298,259 participants from the UK Biobank (Zhang et al., 2023). Their results indicated that nervousness consistently had negative effects on brain health, while warmth, diligence, sociability, and curiosity had protective effects (Zhang et al., 2023). Additional neuroimaging analyses highlighted the associations between personality traits and the frontal cortex, temporal cortex, and thalamus. The study also identified the roles of inflammation and lipid metabolism in the links between personality and brain health (Zhang et al., 2023).

In a study by Graham and collaborators who analyzed data from 15 international datasets, representing 44,094 participants, high neuroticism and low conscientiousness, extraversion, and agreeableness were consistent predictors of mortality across studies (Graham et al., 2017). Regarding mortality, no association was found for stroke-related mortality in one early report (Shipley et al., 2007), while another larger study found a higher risk with higher extraversion and lower conscientiousness (Jokela et al., 2014). Additionally, conscientious individuals are more likely to engage in behaviors that promote positive health and also refrain from risky behaviors that make poorer health outcomes more likely (Wright et al., 2022). Unfortunately, no interventions have been reported in secondary stroke prevention using this data, so potential implications remain speculative.

Conclusion

Current evidence suggests that higher levels of neuroticism and lower levels of conscientiousness may be novel vascular risk factors, increasing the risk of stroke, vascular brain damage, and other health conditions. These associations may be partly mediated by known vascular risk factors, as well as by various, still largely unknown, inflammatory and circulatory mechanisms. Unfortunately, it remains unclear how these findings translate to everyday clinical practice, highlighting the need for additional research. Future studies involving larger and more diverse participant groups with extended prospective follow-ups are essential to better understand the relationship between various personality traits and both brain and overall health. It is also important to investigate other potential confounding factors from both genetic and epigenetic perspectives and to explore potential interventions.

References

Adachi, T., Tsunekawa, Y., Matsuoka, A., & Tanimura, D. (2021). Association between Big Five Personality Traits and Participation in Cardiac Rehabilitation in Japanese Patients with Cardiovascular Disease: A Retrospective Cohort Study. *International journal of environmental research and public health*, *18*(16), 8589. https://doi.org/10.3390/ijerph18168589

Adachi, T., Tsunekawa, Y., & Tanimura, D. (2022). Association between the Big Five personality traits and medication adherence in patients with cardiovascular disease: A cross-sectional study. *PloS one*, 17(12), e0278534. https://doi.org/10.1371/journal.pone.0278534

Agnafors, S., Norman Kjellström, A., Torgerson, J., & Rusner, M. (2019). Somatic comorbidity in children and adolescents with psychiatric disorders. *European child & adolescent psychiatry*, *28*(11), 1517–1525. https://doi.org/10.1007/s00787–019–01313–9

Aschwanden, D., Strickhouser, J. E., Luchetti, M., Stephan, Y., Sutin, A. R., & Terracciano, A. (2021). Is personality associated with dementia risk? A meta-analytic investigation. *Ageing research reviews*, *67*, 101269. https://doi.org/10.1016/j.arr.2021.101269

Bailey R. R. (2016). Lifestyle Modification for Secondary Stroke Prevention. *American journal of lifestyle medicine*, 12(2), 140–147. https://doi.org/10.1177/1559827616633683

Boehme, A. K., Esenwa, C., & Elkind, M. S. (2017). Stroke Risk Factors, Genetics, and Prevention. *Circulation research*, 120(3), 472–495. https://doi.org/10.1161/CIRCRESAHA.116.308398

Bogousslavsky, J. (2003). William Feinberg lecture 2002: Emotions, mood, and behavior after stroke. *Stroke*, 34, 1046–1050

Bruce, J. M., Hancock, L. M., Arnett, P., & Lynch, S. (2010). Treatment adherence in multiple sclerosis: association with emotional status, personality, and cognition. *Journal of behavioral medicine*, *33*(3), 219–227. https://doi.org/10.1007/s10865–010–9247-y

Costa, P. T., Jr, McCrae, R. R., & Löckenhoff, C. E. (2019). Personality Across the Life Span. *Annual review of psychology*, *70*, 423–448. https://doi.org/10.1146/annurev-psych-010418–103244

Cordonnier, C., Sprigg, N., Sandset, E. C., Pavlovic, A., Sunnerhagen, K. S., Caso, V., Christensen, H., & Women Initiative for Stroke in Europe (WISE) group (2017). Stroke in women – from evidence to inequalities. *Nature reviews. Neurology*, *13*(9), 521–532. https://doi.org/10.1038/nrneurol.2017.95

Craske, M.G., Herzallah, M.M., Nusslock, R. *et al.* (2023). From neural circuits to communities: an integrative multidisciplinary roadmap for global mental health. *Nat. Mental Health* **1**, 12–24. https://doi.org/10.1038/s44220–022–00012-w

Čukić, I., & Bates, T. C. (2015). The Association between Neuroticism and Heart Rate Variability Is Not Fully Explained by Cardiovascular Disease and Depression. *PloS one*, *10*(5), e0125882. https://doi.org/10.1371/journal.pone.0125882

DeSoto, M. C., & Salinas, M. (2015). Neuroticism and cortisol: The importance of checking for sex differences. *Psychoneuroendocrinology*, *62*, 174–179. https://doi.org/10.1016/j.psyneuen.2015.07.608

GBD 2019 Stroke Collaborators (2021). Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global

Burden of Disease Study 2019. *The Lancet. Neurology*, *20*(10), 795–820. https://doi.org/10.1016/S1474–4422(21)00252–0

Gil-Salcedo, A., Dugravot, A., Fayosse, A., Landré, B., Jacob, L., Bloomberg, M., Sabia, S., & Schnitzler, A. (2022). Pre-stroke Disability and Long-Term Functional Limitations in Stroke Survivors: Findings From More of 12 Years of Follow-Up Across Three International Surveys of Aging. *Frontiers in neurology*, *13*, 888119. https://doi.org/10.3389/fneur.2022.888119

Graham EK, Bastarache E, Milad E, Turiano N, Cotter K, & Mroczek D. (2018). Physical Activity Mediates the Association between Personality and Biomarkers of Health. *SAGE Open Medicine*, 6, 1–10. 10.1177/2050312118774990

Graham, E. K., Rutsohn, J. P., Turiano, N. A., Bendayan, R., Batterham, P. J., Gerstorf, D., Katz, M. J., Reynolds, C. A., Sharp, E. S., Yoneda, T. B., Bastarache, E. D., Elleman, L. G., Zelinski, E. M., Johansson, B., Kuh, D., Barnes, L. L., Bennett, D. A., Deeg, D. J. H., Lipton, R. B., Pedersen, N. L., ... Mroczek, D. K. (2017). Personality Predicts Mortality Risk: An Integrative Data Analysis of 15 International Longitudinal Studies. *Journal of research in personality*, *70*, 174–186. https://doi.org/10.1016/j.jrp.2017.07.005

Greve, K. W., Sherwin, E., Stanford, M. S., Mathias, C., Love, J., & Ramzinski, P. (2001). Personality and neurocognitive correlates of impulsive aggression in long-term survivors of severe traumatic brain injury. *Brain injury*, *15*(3), 255–262. https://doi.org/10.1080/026990501300005695

Hakulinen, C., Hintsanen, M., Munafò, M. R., Virtanen, M., Kivimäki, M., Batty, G. D., & Jokela, M. (2015). Personality and smoking: individual-participant meta-analysis of nine cohort studies. *Addiction (Abingdon, England)*, *110*(11), 1844–1852. https://doi.org/10.1111/add.13079

Halalmeh, D. R., Salama, H. Z., LeUnes, E., Feitosa, D., Ansari, Y., Sachwani-Daswani, G. R., & Moisi, M. D. (2024). The Role of Neuropsychology in Traumatic Brain Injury: Comprehensive Literature Review. *World neurosurgery*, 183, 128–143. https://doi.org/10.1016/j.wneu.2023.12.069

Ibanez, A., & Zimmer, E. R. (2023). Time to synergize mental health with brain health. *Nature. Mental health*, *1*(7), 441–443. https://doi.org/10.1038/s44220–023–00086–0

Jackson, J. J., Connolly, J. J., Garrison, S. M., Leveille, M. M., & Connolly, S. L. (2015). Your friends know how long you will live: a 75-year study of peer-rated personality traits. *Psychological science*, *26*(3), 335–340. https://doi.org/10.1177/0956797614561800

Jayawickreme, E., Infurna, F. J., Alajak, K., Blackie, L. E. R., Chopik, W. J., Chung, J. M., Dorfman, A., Fleeson, W., Forgeard, M. J. C., Frazier, P., Furr, R. M., Grossmann, I., Heller, A. S., Laceulle, O. M., Lucas, R. E., Luhmann, M., Luong, G., Meijer, L., McLean, K. C., Park, C. L., ... Zonneveld, R. (2021). Post-traumatic growth as positive personality change: Challenges, opportunities, and recommendations. *Journal of personality*, 89(1), 145–165. https://doi.org/10.1111/jopy.12591

Jokela, M., Pulkki-Råback, L., Elovainio, M., & Kivimäki, M. (2014). Personality traits as risk factors for stroke and coronary heart disease mortality: pooled analysis of three cohort studies. *Journal of behavioral medicine*, *37*(5), 881–889. https://doi.org/10.1007/s10865–013–9548-z

Kim J. S. (2016). Post-stroke Mood and Emotional Disturbances: Pharmacological Therapy Based on Mechanisms. *Journal of stroke*, *18*(3), 244–255. https://doi.org/10.5853/jos.2016.01144

Ko, K. M., Moon, S. J., Koh, J. H., Pae, C. U., & Min, J. K. (2020). Contribution of Personality Traits, Psychological Factors, and Health-Related Quality of Life to Medication Adherence in Patients with Rheumatoid Arthritis. *Yonsei medical journal*, 61(5), 406–415. https://doi.org/10.3349/ymj.2020.61.5.406

Kohli RK. (2017). A Systematic Review To Evaluate The Association Between Medication Adherence And Personality Traits. *Value in Health*, 20(9), PA686. DOI:https://doi.org/10.1016/j.jval.2017.08.1732

Kuźma, E., Lourida, I., Moore, S. F., Levine, D. A., Ukoumunne, O. C., & Llewellyn, D. J. (2018). Stroke and dementia risk: A systematic review and meta-analysis. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, *14*(11), 1416–1426. https://doi.org/10.1016/j.jalz.2018.06.3061

Larsen R.J., Buss D.M. Personality psychology: domains of knowledge about human nature. New York (NY): McGraw Hill Education; 2017

Leger, K. A., Turiano, N. A., Bowling, W., Burris, J. L., & Almeida, D. M. (2021). Personality Traits Predict Long-Term Physical Health via Affect Reactivity to Daily Stressors. *Psychological science*, *32*(5), 755–765. https://doi.org/10.1177/0956797620980738

Leon, L., Redondo, M., Garcia-Vadillo, A., Perez-Nieto, M. A., Rodriguez-Rodriguez, L., Jover, J. A., Gonzalez-Alvaro, I., & Abasolo, L. (2016). Influence of patient personality in the treatment of rheumatoid arthritis. *Rheumatology international*, *36*(11), 1549–1555. https://doi.org/10.1007/s00296–016–3561–9

Levine, D. A., Galecki, A. T., Langa, K. M., Unverzagt, F. W., Kabeto, M. U., Giordani, B., & Wadley, V. G. (2015). Trajectory of Cognitive Decline After Incident Stroke. *JAMA*, *314*(1), 41–51. https://doi.org/10.1001/jama.2015.6968

Lo Buono, V., Noto, F., Bonanno, L., Formica, C., & Corallo, F. (2022). Investigations of Personality Trait in Subacute Post-Stroke Patients: Some Preliminary Observations. *Medicina (Kaunas, Lithuania)*, 58(5), 683. https://doi.org/10.3390/medicina58050683

Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: longitudinal analyses with a national sample from Germany. *Journal of personality and social psychology*, *101*(4), 847–861. https://doi.org/10.1037/a0024298

Ma, Z., He, W., Zhou, Y., Mai, L., Xu, L., Li, C., & Li, M. (2024). Global burden of stroke in adolescents and young adults (aged 15–39 years) from 1990 to 2019: a comprehensive trend analysis based on the global burden of disease study 2019. *BMC public health*, *24*(1), 2042. https://doi.org/10.1186/s12889–024–19551–1

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467–6494.1992.tb00970.x

Mijajlović, M. D., Pavlović, A., Brainin, M., Heiss, W. D., Quinn, T. J., Ihle-Hansen, H. B., Hermann, D. M., Assayag, E. B., Richard, E., Thiel, A., Kliper, E., Shin, Y. I., Kim, Y. H., Choi, S., Jung, S., Lee, Y. B., Sinanović, O., Levine, D. A., Schlesinger, I., Mead, G., ... Bornstein, N. M. (2017). Post-stroke dementia – a comprehensive review. *BMC medicine*, *15*(1), 11. https://doi.org/10.1186/s12916–017–0779–7

Molloy, G. J., O'Carroll, R. E., & Ferguson, E. (2014). Conscientiousness and medication adherence: a meta-analysis. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine*, *47*(1), 92–101. https://doi.org/10.1007/s12160–013–9524–4

Murphy, S. J., & Werring, D. J. (2020). Stroke: causes and clinical features. *Medicine (Abingdon, England: UK ed.)*, *48*(9), 561–566. https://doi.org/10.1016/j.mpmed.2020.06.002

Niewada, M., & Michel, P. (2016). Lifestyle modification for stroke prevention: facts and fiction. *Current opinion in neurology*, 29(1), 9–13. https://doi.org/10.1097/WCO.0000000000000285

O'Cleirigh, C., Ironson, G., Weiss, A., & Costa, P. T., Jr (2007). Conscientiousness predicts disease progression (CD4 number and viral load) in people living with HIV. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association*, *26*(4), 473–480. https://doi.org/10.1037/0278–6133.26.4.473

Osterberg, L., & Blaschke, T. (2005). Adherence to medication. *The New England journal of medicine*, *353*(5), 487–497. https://doi.org/10.1056/NEJMra050100

Pavlovic A. M. (2021). Cerebral small vessel disease and heart rate variability: A quest for nontraditional risk factors. *Journal of clinical hypertension (Greenwich, Conn.),* 23(9), 1810–1812. https://doi.org/10.1111/jch.14350

Pavlovic, A. M., Pekmezovic, T., Zidverc Trajkovic, J., Svabic Medjedovic, T., Veselinovic, N., Radojicic, A., Mijajlovic, M., Tomic, G., Jovanovic, Z., Norton, M., & Sternic, N. (2016). Baseline characteristic of patients presenting with lacunar stroke and cerebral small vessel disease may predict future development of depression. *International journal of geriatric psychiatry*, *31*(1), 58–65. https://doi.org/10.1002/gps.4289

Pendlebury, S. T., & Rothwell, P. M. (2009). Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. *The Lancet. Neurology*, *8*(11), 1006–1018. https://doi.org/10.1016/S1474–4422(09)70236–4

Podell, K., Gifford, K., Bougakov, D., & Goldberg, E. (2010). Neuropsychological assessment in traumatic brain injury. *The Psychiatric clinics of North America*, *33*(4), 855–876. https://doi.org/10.1016/j.psc.2010.08.003

Rieger SM. (2015). Personality and Behavior Changes Subsequent to Traumatic Brain Injury: A Review of the Literature. *International Journal of Emergency Mental Health and Human Resilience*, Vol. 17, No.2, pp. 469–474, ISSN 1522–4821

Santos, E. F., Duarte, C. M., Ferreira, R. O., Pinto, A. M., Geenen, R., & da Silva, J. P. (2019). Multifactorial explanatory model of depression in patients with rheumatoid arthritis: a structural equation approach. *Clinical and experimental rheumatology*, *37*(4), 641–648

Sennfält, S., Norrving, B., Petersson, J., & Ullberg, T. (2019). Long-Term Survival and Function After Stroke: A Longitudinal Observational Study From the Swedish Stroke Register. *Stroke*, *50*(1), 53–61. https://doi.org/10.1161/STROKEAHA.118.022913

Shipley, B. A., Weiss, A., Der, G., Taylor, M. D., & Deary, I. J. (2007). Neuroticism, extraversion, and mortality in the UK Health and Lifestyle Survey: a 21-year prospective cohort study. *Psychosomatic medicine*, *69*(9), 923–931. https://doi.org/10.1097/PSY.0b013e31815abf83

Soliemanifar, O., Soleymanifar, A., & Afrisham, R. (2018). Relationship between Personality and Biological Reactivity to Stress: A Review. *Psychiatry investigation*, *15*(12), 1100–1114. https://doi.org/10.30773/pi.2018.10.14.2

Stephan, Y., Sutin, A. R., Luchetti, M., Aschwanden, D., & Terracciano, A. (2023). Personality and Risk of Incident Stroke in 6 Prospective Studies. *Stroke*, 54(8), 2069–2076. https://doi.org/10.1161/STROKEAHA.123.042617

Stilley, C. S., Sereika, S., Muldoon, M. F., Ryan, C. M., & Dunbar-Jacob, J. (2004). Psychological and cognitive function: predictors of adherence with cholesterol lowering treatment. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine*, *27*(2), 117–124. https://doi.org/10.1207/s15324796abm2702_6

Stone, J., Townend, E., Kwan, J., Haga, K., Dennis, M. S., & Sharpe, M. (2004). Personality change after stroke: some preliminary observations. *Journal of neurology, neurosurgery, and psychiatry*, *75*(12), 1708–1713. https://doi.org/10.1136/jnnp.2004.037887

Sutin, A. R., Brown, J., Luchetti, M., Aschwanden, D., Stephan, Y., & Terracciano, A. (2023). Five-Factor Model Personality Traits and the Trajectory of Episodic Memory: Individual-Participant Meta-Analysis of 471,821 Memory Assessments from 120,640 Participants. *The journals of gerontology. Series B, Psychological sciences and social sciences*, *78*(3), 421–433. https://doi.org/10.1093/geronb/gbac154

Sutin, A. R., Stephan, Y., Luchetti, M., Artese, A., Oshio, A., & Terracciano, A. (2016). The Five-Factor Model of Personality and Physical Inactivity: A Meta-Analysis of 16 Samples. *Journal of research in personality*, *63*, 22–28. https://doi.org/10.1016/j.jrp.2016.05.001

Sutin, A. R., & Terracciano, A. (2016). Personality traits and body mass index: Modifiers and mechanisms. *Psychology & health*, *31*(3), 259–275. https://doi.org/10.1080/08870446.2015.1082561

Sutin, A. R., Terracciano, A., Deiana, B., Naitza, S., Ferrucci, L., Uda, M., Schlessinger, D., & Costa, P. T., Jr (2010). High neuroticism and low conscientiousness are associated with interleukin-6. *Psychological medicine*, *40*(9), 1485–1493. https://doi.org/10.1017/S0033291709992029

Terracciano, A., Aschwanden, D., Passamonti, L., Toschi, N., Stephan, Y., Luchetti, M., Lee, J. H., Sesker, A., O'Súilleabháin, P. S., & Sutin, A. R. (2021a). Is neuroticism differentially associated with risk of Alzheimer's disease, vascular dementia, and frontotemporal dementia?. *Journal of psychiatric research*, *138*, 34–40. https://doi.org/10.1016/j.jpsychires.2021.03.039

Terracciano, A., Aschwanden, D., Stephan, Y., Cerasa, A., Passamonti, L., Toschi, N., & Sutin, A. R. (2021b). Neuroticism and Risk of Parkinson's Disease: A Meta-Analysis. *Movement disorders: official journal of the Movement Disorder Society*, *36*(8), 1863–1870. https://doi.org/10.1002/mds.28575

Terracciano, A., Cenatus, B., Zhu, X., Karakose, S., Stephan, Y., Marcolini, S., De Deyn, P. P., Luchetti, M., & Sutin, A. R. (2023). Neuroticism and white matter hyperintensities. Journal of psychiatric research, 165, 174–179. https://doi.org/10.1016/j.jpsychires.2023.07.026

Thibaut F. (2018). Psychiatric disorders: neurodevelopmental disorders, neurodegenerative disorders, or both? *Dialogues in clinical neuroscience*, 20(4), 251–252. https://doi.org/10.31887/DCNS.2018.20.4/fthibaut

Thomas, M. C., Duggan, K. A., Kamarck, T. W., Wright, A. G. C., Muldoon, M. F., & Manuck, S. B. (2022). Conscientiousness and Cardiometabolic Risk: A Test of the Health Behavior Model of Personality Using Structural Equation Modeling. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine*, *56*(1), 100–111. https://doi.org/10.1093/abm/kaab027

Turiano NA, Mroczek DK, Moynihan J, & Chapman BP (2013). Big 5 personality traits and interleukin-6: Evidence for "healthy Neuroticism" in a US population sample. *Brain, Behavior, and Immunity*, 28, 83–89. 10.1016/j.bbi.2012.10.020

Turiano, N. A., Spiro, A., 3rd, & Mroczek, D. K. (2012). Openness to experience and mortality in men: analysis of trait and facets. *Journal of aging and health*, *24*(4), 654–672. https://doi.org/10.1177/0898264311431303

Verdelho, A., Wardlaw, J., Pavlovic, A., Pantoni, L., Godefroy, O., Duering, M., Charidimou, A., Chabriat, H., & Biessels, G. J. (2021). Cognitive impairment in patients with cerebrovascular disease: A white paper from the links between stroke ESO Dementia Committee. *European stroke journal*, 6(1), 5–17. https://doi.org/10.1177/23969873211000258

Waszczuk, M. A., Li, K., Ruggero, C. J., Clouston, S. A. P., Luft, B. J., & Kotov, R. (2018). Maladaptive Personality Traits and 10-Year Course of Psychiatric and Medical Symptoms and Functional Impairment Following Trauma. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine*, *52*(8), 697–712. https://doi.org/10.1093/abm/kax030

Weston, S. J., Graham, E. K., Turiano, N. A., Aschwanden, D., Booth, T., Harrison, F., James, B. D., Lewis, N. A., Makkar, S. R., Mueller, S., Wisniewski, K. M., Yoneda, T., Zhaoyang, R., Spiro, A., Drewelies, J., Wagner, G. G., Steinhagen-Thiessen, E., Demuth, I., Willis, S., Schaie, K. W., ... Mroczek, D. K. (2020). Is Healthy Neuroticism Associated with Chronic Conditions? A Coordinated Integrative Data Analysis. *Collabra. Psychology*, *6*(1), 42. https://doi.org/10.1525/collabra.267

Weston, S. J., Hill, P. L., & Jackson, J. J. (2015). Personality Traits Predict the Onset of Disease. *Social Psychological and Personality Science*, *6*(3), 309–317. https://doi.org/10.1177/1948550614553248

Wheeler, K., Wagaman, A., & McCord, D. (2012). Personality traits as predictors of adherence in adolescents with type I diabetes. *Journal of child and adolescent psychiatric nursing: official publication of the Association of Child and Adolescent Psychiatric Nurses, Inc*, *25*(2), 66–74. https://doi.org/10.1111/j.1744–6171.2012.00329.x

Wiebe, J. S., & Christensen, A. J. (1997). Health beliefs, personality, and adherence in hemodialysis patients: an interactional perspective. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine*, *19*(1), 30–35. https://doi.org/10.1007/BF02883424

Wright, A. J., Weston, S. J., Norton, S., Voss, M., Bogdan, R., Oltmanns, T. F., & Jackson, J. J. (2022). Prospective self– and informant-personality associations with inflammation, health behaviors, and health indicators. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association*, *41*(2), 121–133. https://doi.org/10.1037/hea0001162

Zhang, YR., Deng, YT., Li, YZ. *et al.* (2023). Personality traits and brain health: a large prospective cohort study. *Nat. Mental Health* 1, 722–735. https://doi.org/10.1038/s44220–023–00119–8

Zhang, S., Xu, M., Liu, Z. J., Feng, J., & Ma, Y. (2020). Neuropsychiatric issues after stroke: Clinical significance and therapeutic implications. *World journal of psychiatry*, 10(6), 125–138. https://doi.org/10.5498/wjp.v10.i6.125

Zhou, J., Fangma, Y., Chen, Z., & Zheng, Y. (2023). Post-Stroke Neuropsychiatric Complications: Types, Pathogenesis, and Therapeutic Intervention. *Aging and disease*, 14(6), 2127–2152. https://doi.org/10.14336/AD.2023.0310–2

Funding

This paper was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (contract number 451-03–66/2024–03/200096 and 451-03–66/2025–03/200096).

https://doi.org/10.18485/baf.2025.2.ch4

Goran Knežević*

MAJOR DISPOSITIONS IN THE PSYCHOLOGY OF INDIVIDUAL DIFFERENCES: CONCEPTUALIZATION, MEASUREMENT, ORIGINS, AND CONSEQUENCES

Abstract: The psychology of individual differences focuses on dispositions—recurrent behavioral, cognitive, or affective tendencies that distinguish one individual from another. But how do we accurately identify and measure these dispositions? What approaches and methodologies ensure that our taxonomies of dispositions are both comprehensive and parsimonious? How do we determine the importance of different dispositions? Given the continual fluctuations in our moods and states, to what extent are dispositional or trait-like constructs useful in understanding and predicting behavior? How can we assess their relevance, especially when considering the influence of powerful situational and social forces on behavior? Even if dispositions demonstrate stability in the short term, do they remain stable throughout the lifespan, or do they vary significantly? What are the origins of these dispositions?

This lecture will explore these questions, with a focus on personality dispositions. The goal is to introduce attendees to conceptual approaches, methodologies, techniques, and the latest advances in the empirical study of personality traits, emphasizing their practical implications and theoretical underpinnings in understanding human behavior.

Key words: Dispositions; Taxonomy of basic/major personality traits; Big Five; HEXACO; Disintegration/Psychoticism

Introduction

Dispositions are enduring tendencies or propensities that guide behavior, emotion, and cognition. There are many types of dispositional constructs in psychology: traits, abilities, aptitudes, attitudes, interests,

* Department of Psychology and LIRA lab, Faculty of Philosophy, University of Belgrade.

values, cognitive styles, and more. In psychology, the term is often used as a synonym for *personality traits*, when referring to stable, broad, cross-culturally universal behavioral tendencies accounting for consistencies in behavior over time and across situations. Traits such as extraversion, emotional stability, and conscientiousness influence a wide range of behaviors across domains, from social functioning to academic performance, from health behavior to political orientation. Their broadness, stability, universality, and predictive power make them central to any comprehensive psychological account of human behavior. The psychology of personality has historically focused on identifying and conceptualizing personality traits, developing instruments for their measurement, and elucidating their biological and environmental origins as well as their psychological, social, and behavioral consequences.

In this article, I elaborate on the conceptual status of major personality traits as psychological dispositions, explore the most influential contemporary taxonomies of traits, touch upon the nature of their temporal and cross-situational consistency, and consider what is currently known about their underlying mechanisms and their broad consequences. The emphasis is on traits as *par excellence* natural psychological constructs, not as metaphysical or philosophical entities, but grounded in empirical research within the field of psychology of individual differences or differential psychology.

Conceptualization of traits

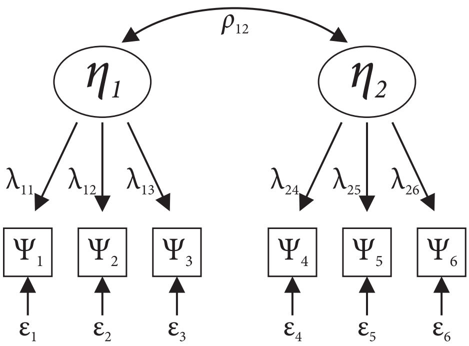
Traits represent an old and intuitive way of understanding human behavior. They are not simple summaries of behavior; rather, traits are latent constructs inferred from the consistency of actions across time and situations. One of the earliest systematic attempts to describe stable personality characteristics can be found in *Characters*, written by Theophrastus, a student of Aristotle, around the 4th century BCE (Theophrastus, 1929). In this work, Theophrastus offered thirty concise sketches of moral and behavioral types, each illustrating a distinct dispositional tendency—such as shamelessness, boorishness, superstitiousness, or meanness. For example, he defined penuriousness (stinginess) as "...excessive economy of expenditure (p. 65)" and listed nineteen specific manifest behaviors indicating this trait in various situations—ranging from reluctance to contribute to public meals to stinginess even in funerary arrangements. This kind of thinking anticipates the modern psychological notion of personality traits as generalized behavioral dispositions manifesting in behavioral regularities in consistent ways over time and across situations.

The idea that individual behavior is governed by such stable dispositions was later challenged during the so-called "personality paradigm crisis" of the late 1960s, spurred in large part by Walter Mischel's influential critique (1968), which questioned the plausibility of traits as scientific constructs. Nonetheless, the trait perspective proved resilient. During the 1980s, it underwent a scientific revitalization and refinement with the emergence of robust empirical models such as the Five-Factor Model (Costa & McCrae, 1992) reinforcing the view that traits are both descriptively useful and empirically defensible constructs for understanding and predicting behavior.

Traits are conceptualized as complex dispositional tendencies inferred from a set of interrelated indicators—such as behaviors, experiences, thoughts, and emotions. They are typically embedded within a broader theoretical framework and are not directly observable. Instead, traits are assessed through multiple observable indicators, and are thus treated as *latent variables*. Core or basic traits, such as extraversion or conscientiousness, are commonly modeled as latent variables—unobserved constructs inferred from systematic patterns in *manifest variables* (Figure 1). Manifest variables (or indicators) are measurable entities that can be directly observed, such as being energetic, lively, sociable, or outgoing—indicators often used to infer extraversion. While manifest variables contain random measurement error, latent variables are mathematically conceptualized as error-free, capturing only the shared, substantive variance across their indicators. This shared variance is assumed to reflect the underlying trait. The extent to which each manifest variable reflects the latent trait is quantified by *factor loadings* (λ), which indicate the strength of the relationship between the manifest and latent variables, or the amount of contribution of latent variables to the variance of manifest variables.

Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) are the primary analytic strategies for examining the structure of latent variables. EFA is aimed to discover structures or patterns among a set of observed variables, used when a researcher has no a priori hypothesis about the structure of the data. CFA (a type of Structural Equation Modeling – SEM) represent the dominant approach to hypothesis testing – it specifies and tests a hypothesized correlational structure or pattern among a set of manifest variables. These models mostly assume *linear relationships* between latent factor and observed variables (the underlying mathematics is based on linear algebra), although there are models that assume *non-linear relationships* between them (e.g., Item Response Theory, IRT, a technique that is mostly used for evaluating test items, assumes logistic relationship between them).

The usual (but not the only one[1]) conceptualization of the relationships between latent and manifest variables is so called *reflective model:* it assumes that the manifest measures result from the latent variable. Reflective models are based on the *"local independence assumption"* – that correlations among the indicators are explained wholly by the latent variable, i.e. correlations observed among the indicators, beyond the latent factor, are spurious. If we consider only persons with the same level on a latent variable, the correlations between the responses are supposed to vanish. The consequence of this is that the indicators in reflective model are not of interest in themselves, they are of interest only as the manifestations of a latent variable, i.e., they are completely interchangeable or replaceable by some other manifest variables demonstrated to be good indicators of the latent variable.



$$y_i = \lambda_{i1}\eta_i + \varepsilon_i \quad \text{for} \quad i = 1, 2, 3$$
$$y_i = \lambda_{i2}\eta_2 + \varepsilon_i \quad \text{for} \quad i = 4, 5, 6$$

Figure 1. η_1, η_2 - latent variables; $\lambda_{11} - \lambda_{26}$ – loadings of the manifest variables on their respective latent factors η_1 and η_2; $\Psi_1 - \Psi_6$ – manifest variables; $\varepsilon_1 - \varepsilon_6$ – error variances; ρ_{12} – correlation between the latent factors.

Broad traits often have multiple dimensions or facets. The set of items in a measure are *unidimensional* when the correlations among the items

1 **One of the alternative conceptualizations gaining prominence in personality psychology is the network model (Cramer et al., 2012).** Unlike traditional latent variable approaches, this model offers a fundamentally different understanding of the relations between traits and their components. In the network approach, components (e.g., symptoms, behaviors, or thoughts) are viewed as causally autonomous—they are not interchangeable or reflective of a single underlying latent trait. Rather, they co-occur because they influence one another directly through causal or logical connections. Despite these different conceptual interpretations stemming from distinct theoretical traditions in different scientific domains, the network and latent variable approaches have been shown to be mathematically equivalent, (Kruis & Maris, 2016).

can be accounted for by a single latent factor. If a trait is *multidimensional*, the correlations among the items can be accounted for by more than one latent factor, reflecting different aspects of this trait. Thus, manifest variables energetic and lively tend to form a subdimenison of Extraversion called Activity, while gregarious and uutgoing tend to form a subdimenison called Sociability. If these different subdimensions reflect the same higher-order trait they have to be substantively interrelated (it should be reflected in a substantial correlation – ρ_{12} – between subdimensions η_1 and η_2, as depticted Figure 1). If a subdimension is not substantively related to the other subdimensions of a trait it means that it indicates some other higher order trait. For example, manifest variables fearful and tense, form a subdimension Anxiety that tend to have low correlations with Activity and Sociability because it is a subdimension of another trait – Neuroticism.

Hierarchical organization of traits

Thus, personality is best conceptualized as a hierarchical structure. At the top of this hierarchy are the broadest traits, followed by their interrelated subdimensions, then more specific and context-sensitive behavioral regularities such as habits, and finally, individual behavioral acts. This organization reflects different levels of analysis of human behavior. As we move upward in the hierarchy, behavioral tendencies become broader, more generalizable across situations, more universal across various human groups, and less dependent on specific situational demands.

The psycholexical approach, initiated by Allport and Odbert (1936) and further developed by Cattell, Fiske, Tupes and Christal, Norman, Goldberg, Saucier and others (John, 2021), provided a naturalistic foundation for identifying core, major, or basic traits through the analysis of natural language—viewed as a cultural repository of accumulated human experience encoded in words. Consequently, an in-depth examination of vocabulary descriptors of human behavior is regarded as a reliable method for ensuring comprehensiveness in the description of human personality. Most psycholexical studies have identified five (John, 2021), six (Ashton et al., 2004), or, in some cases, seven (Saucier, 2003) broad basic traits.

The most robust empirical evidence regarding personality structure – including support based on psycholexical analysis – has been obtained at the highest level of this hierarchy—that of major/basic/core personality traits. These models are, therefore, top-down models. Traits at the lower level of hierarchy of the most influential taxonomies are less robust and less empirically sound, usually not based on factor analysis but conceptual considerations and to an extent arbitrary preference of the authors (e.g., Costa & McCrae, 1995). Basic traits are found to be cross-culturally universal (e.g., McCrae & Costa, 1997), longitudinally stable (e.g., Roberts & DelVecchio, 2000), and heritable (approximately 40% of phenotypic variance is heritable, Vukasović & Bratko, 2015), with identifiable neurobiological underpinnings (e.g., DeYoung et al., 2010). Moreover, basic personality traits are found even across species, e.g., Big Five traits are extracted on ratings of behavior of chimpanzees (Gosling, 2001).

The traditional distinction between temperamental traits (regarded as biologically based, early-appearing, and stable) and personality traits (considered socially shaped and emerging later) is no longer considered viable. Contemporary research demonstrates that all traits arise from complex interactions between genetic/ biological and environmental/ situational factors across the lifespan. Technically, the phenotypic variance of any trait can be decomposed into several components: genetic variance (G), environmental variance (E), the interaction between genetic and environmental factors (G × E), and the correlational effects between genes and environment (doubled G–E correlation). These correlational effects capture the tendency of genetic predispositions to influence the kinds of environments individuals experience, thereby amplifying the expression of certain traits.

What proves to be far more important is the extent to which a trait is, for example, coherent, universal or incrementally relevant. For a behavioral tendency to be recognized as a genuine and fundamental disposition—rather than merely a label for an arbitrary or just conceptually convenient aggregation of loosely connected indicators—it must satisfy several critical criteria.

Criteria for the existence of a trait

Descriptive breadth: The trait should account for a wide range of affective, cognitive, and behavioral tendencies. **Structural independence**: It should be empirically distinguishable (usually through factor analysis) from other traits. **Identification/ extraction across methods**: It should be identified across different methods of assessment, self-report measures, ratings by close others, peer-ratings or professional ratings. **Identification/ extraction across various groups/ populations**: It should be identified in various human subgroups (e.g., subgroups based on gender, age, educational level, nation, culture, or language). **Temporal stability**: Traits exhibit high consistency across time. **Biological basis**: A credible link to genetic, neurobiological, or physiological processes. **Predictive validity**: The ability to predict important life outcomes such as mental health, interpersonal functioning, or societal involvement, incrementally to other traits. **Relevance**: Lexical analysis was the major paradigm indicating relevance of a trait.

This tentative list of criteria is rarely explicated in the literature, but majority of scholars would use these criteria for the existence of a basic dispositional tendency. Majority of these criteria are integrated in so called psychometric evaluation (such as reliability and validity) of the instruments measuring these dispositions. There is persuasive evidence that the basic traits like those captured in the Five-Factor Model (FFM), HEXACO, and Disintegration models fulfill these requirements.

Trait Taxonomies: The Five-Factor Model and HEXACO

The most widely accepted taxonomy of basic/major/core traits today are FFM, also known as the Big Five and HEXACO – a six factor personality model. The FFM encompasses: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These dimensions have been replicated across languages and cultures and show stability across the lifespan (McCrae & Costa, 1997). More recently, the HEXACO model has been proposed by Ashton and Lee (2007), basically preserving Big Five (with some modifications especially regarding the definition of Neuroticism and Agreeableness), but adding a sixth dimension: Honesty-Humility. This trait captures aspects of morality and modesty not well represented in the FFM. Its inclusion is supported by cross-cultural lexical research and by its predictive utility in areas such as unethical and procriminal behavior (Ashton & Lee, 2008). The most recent suggestion regarding the taxonomy of basic personality trait is to upgrade it with one more, i.e., seventh personality dimension – Disintegration/Psychoticism – reflecting the proneness to psychotic-like experiences and behaviors (Knežević et al., 2017; 2022). The predictive validity of this behavioral tendency – not represented by the current personality models – is crucial in areas where irrational beliefs, extreme worldviews and non-analytical thinking styles are present (Knežević et al., 2023; 2024).

Despite differences in structure (number of the domain traits, number and content of their subdimensions), personality models reflect attempts to identify a comprehensive, non-overlapping set of basic personality traits. Here, a taxonomy of seven basic traits will be presented.

Traits and states

The available evidence suggests that although behavior varies across situations, sometimes dramatically, this variation is not random. Instead, traits manifest in behavioral patterns that emerge over time and across settings (Epstein, 1979; 1980).

Trait expression is probabilistic, not deterministic: A highly extraverted person may not act sociably in every context, but across time and situations, such individuals are more likely to seek social engagement. Evidence shows that both highly extraverted and highly introverted would experience all range of extraverted and introverted urges, thoughts and feelings as momentary states. The difference is in the frequency of such states: extraverted persons would experience extraverted urges, thoughts and feelings more frequently across time and situations than introverted individuals. Aggregated over many instances, the pattern becomes clear.

Based on the evidence on how traits manifest as temporary states in given contexts, Fleeson defined "traits as density distribution of states" (2001, p. 1011). In other words, what we define as a trait value is a value close to the means of the distributions of states a person experience in various contexts over time.

Advantages of Trait-Based Approaches

Trait-based approaches offer several key advantages. They provide a unifying framework for understanding individual differences, serve as building blocks for more complex psychological theories, and enable cumulative scientific progress. Traits are relatively easy to measure, show consistency over time and across cultures, and often outperform state-based or situational variables in predicting real-world outcomes. Thus, Disintegration/Psychoticism trait was shown to be a predictor of the proneness to COVID-19 conspiratorial beliefs far stronger then socio-demographics, COVID-19 related experiences, anxiety and depression or interactions between these experiences and related anxious and depressive symptoms, Knežević et al., 2023).

Moreover, traits provide a vocabulary for describing relevant psychological variation in both clinical and non-clinical populations. They also serve as useful heuristics in different applied contexts such as clinical diagnosis (Malouff, et al., 2005) and treatment planning (Bagby et al., 2016), personnel selection and job performance (Ones et al., 2007), education (Poropat, 2009), well-being (Soto, 2021), somatic health and mortality (Roberts et al., 2007) and forensic assessment (Costanzo & Costanzo, 2020).

Basic personality traits: Definition, Structure, and Consequences

Neuroticism/Negative Affectivity/Emotionality

Neuroticism refers to a broad disposition to experience negative emotions such as anxiety, sadness, irritability, and vulnerability. Individuals high in Neuroticism are more likely to perceive situations as threatening, to respond with distress, and to have difficulties in regulating emotional responses. The structure of Neuroticism – like other traits – is multifaceted. In FFM it consists of Anxiety (proneness to worry, fear, and tension), Depression (tendency toward sadness and hopelessness), Self-consciousness (susceptibility to embarrassment and social anxiety), Impulsivity (difficulty in controlling cravings and urges), Angry Hostility (tendency to become easily annoyed or angered) and Vulnerability (difficulty coping with stress or recovering from adversity). In HEXACO model this trait – called Emotionality – is less saturated with irritability and oriented more toward Sentimentality (excessive emotional attachments and an empathic sensitivity to the feelings of others), and Dependability (need for emotional support, seeking for encouragement and comfort), beyond what represent a usual indicators of this trait: Anxiety and Fearfulness (tendency to experience fear, avoidance of physical harm).

These facets are moderately intercorrelated and together reflect a general sensitivity to negative affect, primarily anxiety and fear. It represents the opposite pole of emotional resilience.

In terms of outcomes, Neuroticism is one of the strongest predictors of psychological distress, including depression, anxiety disorders, and general dissatisfaction with life. It also shows consistent associations with poor physical health, increased stress reactivity, and maladaptive coping styles. Moreover, high levels of Neuroticism are linked to interpersonal difficulties, such as relationship conflicts and lower social support, as well as occupational outcomes like job dissatisfaction and burnout.

Despite its negative connotations, Neuroticism may have adaptive aspects in specific contexts. For example, heightened vigilance and sensitivity to threat might serve as protective functions in dangerous environments. Nonetheless, the overall effect of high Neuroticism is typically unfavorable, especially when chronic and unbuffered by protective traits or coping mechanisms.

Extraversion/Positive affectivity

Extraversion is a fundamental personality trait reflecting the degree to which individuals are oriented toward the external world. Extraverted individuals are typically energetic, sociable, assertive, and experience positive emotions more frequently. In contrast, introverted individuals are more reserved, reflective, and often prefer solitary activities. In FFM, core facets of Extraversion include Sociability (desire for and enjoyment of social interaction), Assertiveness (tendency to speak up, take charge, and lead), Warmth (kindness and warmness), Activity (preference for a fast-paced, busy lifestyle), Excitement seeking (pursuit of stimulation and novelty), and Positive Emotions (tendency to experience joy, enthusiasm, and optimism). In HEXACO the facets are Sociability again, then Social Self-Esteem (positive self-regard particularly in social contexts), Liveliness (enthusiasm and energy), and Social Boldness (confidence within a variety of social situations, such as approaching strangers and speaking up in group settings).

Extraversion is strongly linked to a range of desirable life outcomes. Extraverts report higher levels of happiness and well-being, are more likely to pursue and maintain social relationships, and tend to succeed in roles that demand interpersonal engagement, such as sales, management, and politics. However, high levels of Extraversion may also be associated with impulsivity, risk-taking, and lower academic performance when not moderated by traits like Conscientiousness.

Openness

Openness to experience captures the extent to which individuals are imaginative, curious, open-minded, and attuned to aesthetics and abstract thinking. This trait encompasses both intellectual engagement and aesthetic sensitivity. FFM facets include: Imagination (capacity for fantasy), Aesthetic sensitivity (deep appreciation for art and beauty), Emotional openness (receptivity to inner feelings and experiences), Intellectual curiosity (desire to explore novel ideas and concepts, enjoyment in abstract thinking), Openness to action (exploration, willingness to try new activities), and Openness to values (liberalism and willingness to challenge traditional beliefs). In HEXACO facets are: Aesthetic appreciation, Inquisitiveness (seeking information and experience with the natural and human world, curiosity about the natural or social sciences), Creativity (innovation, original thinking), and Unconventionality (preference for novelty and unconventional thinking).

High Openness is related to creativity, tolerance for ambiguity, and preference for complexity and change. It is consistently associated with intellectual achievement, political liberalism, and engagement in artistic and cultural pursuits. On the downside, very high Openness may relate to susceptibility to unrealistic beliefs or unconventional thinking that drifts toward impracticality.

Agreeableness/Friendliness

Agreeableness reflects the degree to which individuals are compassionate, cooperative, and motivated to maintain positive social relationships. High-Agreeableness individuals are perceived as kind, empathetic, trusting, and forgiving. Key FFM facets include: Trust (belief in the sincerity and good intentions of others), Altruism (tendency to help and support others), Compliance (willingness to yield and avoid conflict), Modesty (humility and lack of self-aggrandizement, Straight-forwardness (frankness in communication, sincerity and transparence), and Tender-mindedness (compassion, empathy, concern for others' well-being). In HEXACO facets are Forgiveness (willingness to feel trust and liking toward those who may have caused one harm), Gentleness (tendency to be mild and lenient in dealings with other people), Flexibility (willingness to compromise and cooperate with others), and Patience (tendency to remain calm rather than to become angry).

Agreeableness predicts a range of prosocial behaviors, including cooperation, volunteering, and reduced aggression. It is important for harmonious relationships, group functioning, and moral conduct. However, extremely high Agreeableness can lead to submissiveness, self-neglect, and difficulty asserting one's own interests.

Conscientiousness/Self-control

Conscientiousness describes the degree of organization, responsibility, and goal-directed behavior. Individuals high in Conscientiousness are disciplined, reliable, and achievement-oriented. FFM facets include: Orderliness (preference for structure and organization), Competence (self-efficacy, confidence in one's ability to accomplish tasks), Dutifulness (sense of moral obligation and reliability), Achievement striving (persistent pursuit of excellence), Self-discipline (capacity for sustained effort and impulse control, and Deliberation (tendency to think carefully before acting). HEXACO facets are Organization (Seeking order, keeping things tidy and structured approach to tasks), Diligence (tendency to work hard, strong "'work ethic"), Perfectionism (being concerned with details, checking carefully for mistakes and potential improvements) and Prudence (tendency to deliberate carefully and to inhibit impulses).

Conscientiousness is one of the most consistent predictors of life success. It correlates strongly with academic and occupational performance, health-promoting behaviors, and longevity. At extreme levels, however, it may be linked to compulsiveness, perfectionism, and rigidity.

Disintegration/Psychoticism

Disintegration refers to a dispositional tendency toward psychotic-like experiences and behaviors. It captures aspects of personality missed by both FFM and HEXACO.

Its components include: Perceptual Distortions (depersonalizations, derealizations, hallucinations), Paranoia (suspiciousness, persecutory ideas, blaming others for personal failures), General Cognitive/Executive Impairment (poor execution, organization, and coordination of thoughts, emotions, motoric, and speech outputs), Somatoform Dysregulation (sensory and motor conversions, somatic delusions, pain insensitivity), Enhanced Awareness (dissociative engagement with stimuli, synesthesia, visual reminiscence), Magical Thinking (irrational beliefs, ideas of reference, paranormal beliefs), Flattened Affect (emotional numbing, indifference toward self and others, diminished emotional reactivity/expression), Apathy/Depression (chronic fatigue, hopelessness, suicidal ideation), and Mania (agitation, excessive optimism, grandiosity). PID-5 Psychoticism (Krueger et al., 2012) was another recent conceptualization of the psychosis-proneness. Unlike Disintegration, claimed to be basic personality trait, i.e., relevant to both normal and maladaptive personality variations, PID-5 Psychoticism is defined as a dimension of maladaptive personality variations.

Disintegration has shown unique predictive value for mental health vulnerability and clinically relevant conditions (such as posttraumatic stress disorder, depression, density of psychotic episodes, and grandiose narcissism). It is related to a range of important dispositional tendencies and behaviors beyond clinical setting such as conspiracy and other irrational beliefs, socio-political attitudes, components of militant extremist mind, close-mindedness and behaviors such as adult emotional crying, or questionable medical practices. Despite its primary relevance for maladaptive behavior, it is argued that the proneness to false positive error – as the major perceptual and cognitive mechanism driving individual differences in Disintegration (Knežević, Kušić et al., 2024) – can be adaptive in the situations where the occasional true detection provides a substantial fitness advantage. In such occasions natural selection may favor pattern detection strategies that err on the side of false positives, minimizing thus the costly harm caused by false negative errors.

There are various perspectives on whether heritable individual differences in personality dimensions have adaptive value (e.g., Nettle, 2006), or whether they merely reflect random, evolutionarily irrelevant fluctuations that should not be regarded as adaptations (Tooby & Cosmides, 1999). Nevertheless, the evidence demonstrating both positive and negative adaptive consequences associated with high levels of each of the seven major traits appears to support a particular mechanism for the persistence of individual differences along personality dimensions (Penke et al., 2007): balanced selection driven by environmental heterogeneity. According to this view, each personality dimension results from trade-offs between different fitness costs and benefits. Crucially, there is no universally optimal value for these trade-offs. Environmental conditions, constantly changing across time and space, may favor high levels of a particular trait under certain circumstances but render the same levels disadvantageous under others. Consequently, genetic diversity in personality traits is expected to be maintained within the population.

Conclusion: Personality Traits as Natural Dispositions

Personality traits are psychological dispositions *par excellence*. They represent relatively enduring, general tendencies that shape the way individuals think, feel, and behave. Supported by robust empirical evidence, measured reliably across time and contexts, and predictive of wide-ranging outcomes, traits occupy a central position in contemporary psychology of individual differences.

Future work must continue to integrate trait models with developmental, cognitive, cultural and biological psychology, while remaining attentive to the complexities of person-situation dynamics and the multiplicity of psychological processes. But the core insight remains: To understand a person, one must begin with their basic personality traits.

References

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1), i–171.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11(2), 150–166.

Ashton, M. C., & Lee, K. (2008). The prediction of honesty–humility-related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality*, 42(5), 1216–1228.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., et al. (2004). A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, 86, 356–366. https://doi.org/10.1037/0022–3514.86.2.356.

Bagby, R., Gralnick, T., Al-Dajani, N., & Uliaszek, A. (2016). The Role of the Five-Factor Model in Personality Assessment and Treatment Planning. *Clinical Psychology: Science and Practice*, 23. 365–381. 10.1111/cpsp.12175.

Costanzo, M. A., & Costanzo, M. L. (2020). Personality and forensic psychology. In B. J. Carducci, C. S. Nave, A. Di Fabio, D. H. Saklofske, & C. Stough (Eds.), *The Wiley encyclopedia of personality and individual differences* (pp. 575–579). John Wiley & Sons. https://doi.org/10.1002/9781119547181.ch360

Costa, P. T., & McCrae, R. R. (1992). NEO personality inventory-revised: Professional manual. Odessa, FL: Psychological Assessment Resources Inc..

Costa, P. T., & McCrae, R. R. (1995). Domains and Facets: Hierarchical Personality Assessment Using the Revised NEO Personality Inventory. *Journal of personality assessment*. 64. 21–50. 10.1207/s15327752jpa6401_2.

Cramer, A. O. J., Van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality, 26*(4), 414–431. https://doi.org/10.1002/per.1866

DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the Big Five. *Psychological Science*, 21(6), 820–828.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37*(7), 1097–1126. https://doi.org/10.1037/0022–3514.37.7.1097

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist, 35*(9), 790–806. https://doi.org/10.1037/0003–066X.35.9.790

Fleeson, W. (2001). Toward a structure– and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80(6), 1011–1027. https://doi.org/10.1037/0022–3514.80.6.1011

Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin*, 127(1), 45–86. doi:10.1037/0033–2909.127.1.45.

Knežević, G., Kušić, M., Lukić, P., Lazarević, L. B., & Keller, J. (2024). Can the "shotgun wedding" of Openness and Psychoticism be justified based on apophenia as the disposition to commit false positive errors? *Zeitschrift für Psychologie*, 232(4), 255–268. https://doi.org/10.1027/2151–2604/a000580

Knežević, G., Lazarević, Lj., Bošnjak, M., & Keller, J. (2022). Proneness to psychotic-like experiences as a basic personality trait complementing the HEXACO model—A preregistered cross-national study. *Personality and Mental Health,* 16, 244–262. https://doi.org/10.1002/pmh.1537

Knežević, G., Lazarević, L., Mihić, L., Pejović-Milovančević, M., Terzić, Z., Tošković, O. et al. (2023). The anatomy of COVID-19-related conspiracy beliefs: Exploring their nomological network on a nationally representative sample. *British Journal of Clinical Psychology, 62, 642–662.* https://doi.org/10.1111/bjc.12431

Knežević, G., Petrović M. B., Ninković, M., Zupan, Z., Lukić, P., Purić, D., et al. (2024). Lifetime prevalence of questionable health behaviors and their psychological roots: A preregistered nationally representative survey. *PLoS ONE* 19(11): e0313173. https://doi.org/10.1371/journal.pone.0313173

Knežević, G., Savić, D., Kutlešić, V., & Opačić, G. (2017). Disintegration: A Reconceptualization of Psychosis Proneness as a Personality Trait Separate from the Big Five. *Journal of Research in Personality,* 70, 187–201. http://dx.doi.org/10.1016/j.jrp.2017.06.001

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait mode. *Psychological Medicine*, 42, 1879–1890.

Kruis, J., Maris, G. (2016). Three representations of the Ising model. *Scientific Reports,* 6, 34175 (2016). https://doi.org/10.1038/srep34175.

Malouff, J. M., Thorsteinsson, E. B., & Schutte, N. S. (2005). The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment*, 27(2), 101–114.

https://doi.org/10.1007/s10862–005–5384-y

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52(5), 509–516.

Mischel, W. (1968). Personality and Assessment. New York: John Wiley & Sons, Inc.

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist*, 61, 622–631. http://dx.doi.org/10.1037/0003–066X.61.6.622.

Ones, D., Dilchert, S. Viswesvaran, C. & Judge, A. T. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*. 60. 995 – 1027. 10.1111/j.1744–6570.2007.00099.x.

Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality*, 21, 549–587.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322–338. https://doi.org/10.1037/a0014996

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126(1), 3–25. https://doi.org/10.1037/0033–2909.126.1.3

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313–345.

Saucier, G. (2003a). An alternative multiple-language structure of personality attributes. *European Journal of Personality*, 17,179–205. http://dx.doi.org/10.1002/per.489.

Soto, C. (2021). Do Links Between Personality and Life Outcomes Generalize? Testing the Robustness of Trait–Outcome Associations Across Gender, Age, Ethnicity, and Analytic Approaches. *Social Psychological and Personality Science*, 12, 118–130. 10.1177/1948550619900572.

Theophrastus (1929). *The Characters of Theophrastus* (J. M. Edmonds, Ed. & Trans.). London: William Heinemann; New York: G. P. Putnam's Sons.

Tooby, J. & Cosmides, L. (1990). On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation. *Journal of Personality*, 58, 18–67.

Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin*, 141(4), 769–785. doi:10.1037/bul0000017.

3. EPISTEMIC DISPOSITIONS

https://doi.org/10.18485/baf.2025.2.ch5

Iris Vidmar Jovanović

EPISTEMIC DISPOSITIONS OF LITERATURE: INSIGHTS FROM LITERARY TRADITION AND CRITICAL PRACTICE

Abstract: My aim in this paper is to explore current philosophical debates on the cognitive value of literary fiction, primarily in light of Gregory Currie's recent anti-cognitivist account. According to Currie, the claim that fiction is a source of knowledge and other cognitively valuable states is a dispositional one: what we mean by it is that fiction has the capacity to inspire positive cognitive change in the audience. For such a claim to be true however, we need to show that someone has in fact learnt something from art, and this is, Curries argues, hard to do for various reasons. Having extensively discussed these reasons, Curries rejects the view that we learn from fiction, suggesting that philosophical belief in epistemic dispositions of literary fiction is misguided and lacks empirical support. In my response, I focus on several elements of Currie's criticism in order to show that his most elaborate claims – the one refuting the epistemic reliability of fiction and the one asking for empirical evidence supporting cognitivists' credo – are not in fact as detrimental to aesthetic cognitivism as he takes them to be.

1. Literature and its cognitive value

In discussing Dostoyevsky's *Crime and Punishment*, Ilham Dilman writes: "The novel is a study of crime and punishment ... of the evil that enters into a soul of the man who consents to kill another human being, and of the way he finds his way back to good through the acceptance of punishment. Dostoyevsky is interested in the conditions that make the soul vulnerable to such an evil." (1984, 97). Having explored what these conditions may be, relating them, among other things, to the particular social circumstances within which the main character finds himself, his acceptance of utilitarianism and socialism, his encompassing feeling of guilt brought about by his mother, Dilman argues: "Dostoyevsky is interested in the way pride, humiliation, anger and resentment, can turn into a force of evil and feed on each other, and in the way they lend their energy to ideas that inspire the desire for grandeur in the self and contempt for other people." (1984, 97). His analysis of this literary masterpiece is firmly rooted in psychological literature, but even those who do not think we should turn to Freud to understand fictional characters can appreciate the extent to which Dilman's reading of the novel sheds light on certain unfathomable acts of brutality. And those with sufficient level of philosophical imagination can quickly recognize the way his analysis feeds into ethical questions lurking behind consequentialist ethics, notions of character and theories of punishment. One can do a lot of things with this novel, but one cannot ignore the tremendous intellectual impact that it has, in igniting our reflective processes in directions as diverse as psychology, philosophy and sociology.

Dilman's account highlights the capacity of literature to reveal to us something about the world, to help us understand how certain conditions came about, what factors gave rise to them, and what consequences followed. What we see here is an example of what I call epistemic dispositions of literature: its capacity to contribute to human epistemic goals of gaining knowledge and understanding of the world and other people, themselves, and the complexities of human life. These kinds of dispositions are also evident in literature's capacity to contribute to the development of our epistemic values, primarily those related to the virtues of the mind, such as open-mindedness, critical thinking and awareness. By confronting us with a character as complex and intriguing as Raskolnikov, Dostoyevsky pushes us to become curious, to reconsider our conceptions of right and wrong, to look at the world from a more informed perspective. Doing so is important also for our processes of moral reasoning which underlie our daily ethical comport with other epistemic and ethical agents. As Nora Hamalainen argues, "literature is conceptualized as a privileged medium for picturing and understanding moral development, because it can make vivid for us what the transition from one moral conceptions or way of looking to another is like in an individual human life, for a whole, idiosyncratic person." (2019, 31).

The task of explaining epistemic dispositions of literature involves showing how engaging with it can offer cognitive benefits, or how readers can, in the process of reading a certain work, cash out the cognitive and ethical potential that a work offers. This task has largely been carried out by aestheticians, more so than by epistemologists or moral philosophers, which is why there is usually more philosophical interest in exploring how a work's cognitive/ethical dimension contributes to its aesthetic or artistic value, than in explaining the epistemological sources of cognitive dimension.[1] Aesthetic cognitivism, a term that encompasses a cluster of views that recognize art's cognitive dimension, posits that the cognitive merits of art, such as the knowledge it can offer, are often intertwined with its aesthetic qualities. I adopt this framework and I approach it from an epistemological perspective, exploring how literature (and narrative art generally) can provide reliable and justified knowledge. I adopt this perspective in order to critically engage with one of the most ferocious challenges to aesthetic cognitivism defined recently, the one articulated by Gregory Currie in his insightful and challenging book, *Imagining and Knowing: The Shape of Fiction*. While Currie raises immensely important questions concerning the epistemic dispositions of literature, my sense is that his skepticism is exaggerated. To show that, I focus on one particular claim he makes in support of his position – the one challenging the institutional foundations of literature's epistemic value – and I show that it is not entirely convincing by drawing on the examples from literary tradition and critical practice.

2. Currie's many challenges

Currie's analysis of the many ways in which aesthetic cognitivists are wrong is by far too complex to address it in a single paper, which is why I focus on one of his claims that strike me as most concerning. First however let me express one crucial point on which I agree with him. I will adopt his framework of speaking about aesthetic cognitivism in terms of literature's disposition to have a particular impact on readers; as stated above, I am interested in dispositions related to epistemic (cognitive) and moral impact that art can have on the audience, in helping them become better epistemic and moral agents. Such impact can include the acquisition of knowledge, development of a new perspective or a point of view, a deeper understanding, a more informed background on certain topics, etc. But as Currie points out, it is not entirely clear what the cognitivists have in mind when they talk about literature's capacity to help readers reach these states: how would we, after all, prove that such changes have taken place? Consider Currie's approach:

---

1 Some exceptions include Nussbaum's (1986), work on the ethical dimension of Greek tragedies and realist novels, Diamond's (2010) account of literature's relation to morality; Young (2001) approach to artworks as representations, Elgin's (2007), and Davies' (2007) analogy between literature and thought experiments; Carroll's (2007) account of realist fiction (which I address below). I defended the epistemic value of literature by drawing the analogy with testimony (2019). For the importance of authorial testimony from the perspective of philosophy of language and debate on fiction, see Stock (2017).

> The claim that the work contains or is at least in some way a source of the know-how [and other epistemic gains] amounts to this: that people are apt (in circumstances that no doubt need further specification) to gain the know-how from exposure to that work. It ascribes, in other words, a disposition to the work. And claims that something is disposed to affect other things in certain ways are empirical and so in need of evidential support... (Currie 2020,100).

What Currie points to here is that we should only accept aesthetic cognitivism if we can in fact prove that readers change in the process or after the process of reading: that they really learn something, that they became more aware of something they did not realize before their exposure to the particular work, that their perspective is wider or better informed, etc. For the most part however, ever since Aristotle's original attempt to ground aesthetic cognitivism,[2] philosophers have ignored the practical, empirical side of the question, with little interest in empirically proving their claims. Within the public discourse however, concerns regarding dispositions of some literary works, and of some films, have repeatedly been voiced and they often played a role in public policies orchestrating the spectators' access to these works – consider the worries over Gustave Flaubert's *Madam Bovary* or Charles Baudelaire's poetry, Nabokov's *Lolita* or contemporary initiatives to remove novels by John Steinbeck or Mark Twain on the account of their alleged racism.[3] Currie is, much like Plato was, sensitive of the importance of the empirical challenge but whereas Plato feared the moral corruption inflicted by literature, Currie's take on it remains theoretical, in not suggesting or advocating practical consequences of his view. But the worry goes in both directions: unless we can prove the actual positive benefits of a given work, we lose at least some of the reasons to insist on the educational importance of the work itself. On the other hand, we also lose the basis on which to restrict spectators' access to it: if we can't prove that reading *Gone with the wind* will trigger or reinforce racists attitudes, a decision made by certain universities to remove the novel from educational curriculum is hard to justify, particularly given the work's aesthetic qualities and literary value.

I fully agree with Currie's emphasis on empirical challenge and I have elsewhere (2021; 2023) supported his call for empirical settlement of the question of epistemic dispositions of literature, with respect to art's alleged positive impact, as well as with respect to its potentially harmful impact. The problem of empirical testing is quite complicated, and here I can't go into details, other than to reverberate Currie's argument: in order to claim

2 See Aristotle's *Poetics*.

3 See Ladenson (2007); Appleman (2021).

anything about literature's epistemic dispositions, we first need to see the kind of beings we end up being after reading – precisely as Currie argues. However, as I argued, much of the empirical studies currently conducted fail to consider epistemic and ethical character of the spectators (the exceptions are studies which look into character/personality traits[4]) and to explore those factors that make them open to the dispositions that any given book offers. Just like documentaries about the climate change fail to inspire change in our behavior towards the environment, so too can a novel about the hardships of immigrants fail to inspire a change in how we treat these people. On my view, that is not the fault of the work, at least not entirely: epistemic dispositions of literature will only be activated with those readers who can not only recognize them, but are willing to act on them.[5] Yet, this is not something that the research is sensitive too, which is why I do not think the empirical research as currently conducted is going to settle the issue of epistemic dispositions of literature.

Additional problem regarding the empirical challenge concerns the indeterminacy of the results and more general issues related to conducting the research in the first place.[6] This is why it is hard to expect the research will explain away either the aesthetic cognitivists' faith in the positive cognitive impact of literature, or the worry of those who see art as potentially corruptive. For this reason, in what follows I rely on the literary tradition and critical practice to refute Currie's denial of the cognitive value of literature. First however let us look in more details one of Currie's insightful arguments that undermine our faith in epistemic dispositions of literature: the one invoking what he calls institutional constraints.

As mentioned, the claim that narrative art is cognitively valuable implies that it offers intellectual benefits to its audience, particularly to active readers who reflect on the work. Through engaging with the narrative, readers can enrich their understanding, develop their imagination, refine their moral sense, and learn new things. Not all narrative works offer these cognitive benefits, but some do, and my paper aims to explore how literary writers achieve this effect. Currie however cautions us to be careful: as he argues, literature per se is not inherently related to knowledge-seeking practices in the way in which science is. In other words,

4 See Kjeldgaard-Christiansen (2021).

5 In claiming this I am somewhat diminishing the power of literature, or, as some would put it, its value. Matthew Kieran (2006) for example ties the value of literature with its capacity to inspire real change. My own view (2023), inspired by Alan Goldman's (2009) account of moral motivation, does not put such pressure on literary works and focuses rather on the spectators' character.

6 McGregor (2018) offers an overview of such issues, arguing, in conclusion, that it is rather unlikely that such research can be conducted.

while the institutions of science constrain activities of scientists in ways that make it more likely that they will come up with well-confirmed, better evidenced theories that amount to the progression of knowledge, no such framework exists with respect to literature: there is not a framework that would constrain epistemic goals of literature. For example, there are no "assessments of the psychological insights that literary works provide" (185). Consequently, there are no epistemic mechanisms that guide either the authorial work, or the readers' engagement. In addition, the outcome of scientific research – scientific works, papers, theories, models, etc. – are delivered by experts and are intended for the audience sufficiently familiar with the relevant background. Consequently, the audience shares the same or sufficiently similar level of expertisim with the science authors, which enables it to evaluate the reliability of the proposed theories and point to errors, inconsistencies and other knowledge-reducing elements in their works. With respect to literature there are no such mechanisms; neither is there a body of experts monitoring what literary authors write, nor are the readers experts with respect to what they are reading. Consequently, it is hard to see why we should take literature as epistemically reliable source of knowledge.

Currie's argument here is not original; we find similar claims in Plato's *Republic*, and, more recently, in Jerome Stolnitz's (1992) famous attack on the cognitive value of art. Nevertheless, Currie draws our attention to an important aspect of our literary engagements in cautioning us to question our readiness to believe what we read. His issue is all the more pressing, considering how often readers form beliefs about the world on the grounds of fictional sources. James Young reports: "In one study, subjects were given a general knowledge questionnaire after reading a work of fiction. The subjects drew on information from the fiction in completing the questionnaire. They often believed that the information they used in answering the questionnaire was possessed prior to reading the work of fiction." (2017, 86). The problem of course is, sometimes works are factually false, but readers may nevertheless rely on them in forming their beliefs. The question before us now is, does that diminish the overall cognitivists' account of the epistemic dispositions of literature? To this we turn next.

3. Epistemic constraints in literature

My starting point in developing an account of the epistemic dispositions of literature is inspired by John Gibson's take on aesthetic cognitivism. As he argues,

> the literary perspective (and the artistic perspective more generally) is the definitive *human* perspective: the standpoint from which we are best able to bring to light the range of values, desires, frustrations, experiences, and practices that define the human situation. On this view, works of literature, at least when they live up to their promise, represent cognitive achievements: they embody ways of *knowing* the world. (Gibson 2009, 467).

The metaphor that Gibson employs in order to account for the way literature provides us with knowledge is that of an 'archive' that comprises our literary heritage to which we can turn to explore the history of our social, political, economic, religious, philosophical and other segments of the world.[7] As he explains:

> it strikes me as plausible to think that the act of telling a story, of weaving a narrative, is a way of giving structure to a certain conception of human experience and circumstances – a way in which we 'are' in the world. And the point I am putting on offer is that writers of many of our great works of literature offer us narratives that (...) 'give order' to the world (...) by being representative of various regions of it (Gibson 2007, 72).

Part of Gibson's argument is the claim that, in writing, "writers often assume the role of documentarian. They try to "bear witness" through their stories, and, to this extent, one of the *narrative goals* of many literary works is to explore practices, institutions, and forms of interaction that shape our world and structure our experience and relationships." (Gibson 2011). He gives the example of Faulkner's novel *Intruder in the Dust*:

> Faulkner's accomplishment was not to construct a terribly original story but to tell a story in a particular way, a way that rendered intelligible how certain features of Southern culture give rise to these familiar, intractable problems of race. Faulkner was a writer of fiction and not a sociologist, so his work consisted not in statistical surveys but in telling a story that reflects in general way how Southern culture hangs together: that reflects its character, at least once upon a time and from one vantage-point. What Gibson emphasizes is literature's ability to present to readers "a feel for the fabric of the culture [that is] a sense, in short, of how the culture it explores is constituted, at least in respect to the questions Faulkner asks of it (Gibson 2011).

If literature can indeed have this function, we need to show that Currie's concerns regarding its epistemic reliability can be addressed. In other words, we need to show that literature is immersed into the real world and that it reflects real world back to its readers in manners which enables them to pick up and cash out on its epistemic dispositions. A good place to start developing such an account is Noël Carroll's claim:

7 See Gibson 2007.

> To be truly beholden to the artwork on its own terms, then actually requires acknowledging the kind of work of art it is. And the kind of artwork it is may mandate that one attend to it as something other than a *sui generis* formal design. (...) Consider, for example, the realist novel. Given the nature of the realist novel, we do not expect of it that it simply confects a coherent novelistic form. The realist novelist is also expected, as part of his *essential* job description, to be an accurate and penetrating observer of society, or, at least, of the social milieu that he describes. This is absolutely central to his charge as a realist novelist (Carroll 2007, p. 32).

A lesson from Carroll is that, at least with respect to the realist novels, we are justified in accepting them as true, *not doing that* would mean that we don't know how to participate in a literary practice. Knowledge of the real world that we bring into reading, as well as knowledge and familiarity with various genre conventions, help us differentiate between what is true and what is not. The implication is not that a reader will know in each particular case what is true and what is not, that is, in which matters the author is reliable, but that doesn't mean that, generally speaking, all authors are unreliable all the time.

Carroll's insistence on realism's relation to the truth is evident in numerous works of literary criticism. Gregg Crane shows how the writers pertaining to the genre "... share a general conception of fiction as a detailed and accurate representation of historically specific characters and settings – their manners, ways of dress, speech patterns, social habits, main concerns, and topics of conversation" (Crane 2007, 156). Crane's analysis of literary realism reveals realism as "empiricist in orientation", grounded upon "concrete examples", focused on an "exploration of the here and now", on the "world of concrete personal experiences", "inductive rather than deductive, experimental and open to uncertainty" (Crane 2007, 157–8). Realism is also characterized by depicting everyday aspects of life:

> In pursuing detailed rendering of the scenes before them, realists were, to varying degrees, determined to examine subjects, images, or actions previously scorned by sentimental novelists and romancers as common, brutal, or even sordid. So in addition to portraying everyday events such as ordering lunch for the first time in a Chicago cafe (*Sister Carrie*) or riding a street car in lower Manhattan (*A Hazard of New Fortunes*), they also describe with striking frankness war, suicide, disease, crime, and poverty. (Crane 2007, 165)

The idea of a realist writer as 'observer' of society can be applied to all the writers pertaining to this genre. In reference to William Dean Howells's novel *A Hazard of New Fortunes*, Crane writes:

> Howells paints a panoramic portrait of urban life. His novel abounds in richly detailed descriptions of people representing the socio-economic spectrum, including recent immigrants, transplanted Southerners, old money and the newly rich, artists and writers. The points of view expressed by these characters include a property-is-theft socialism, a conservative Gospel of Wealth capitalism, and a remnant of the Old South's feudal aristocratic perspective. The crisis of Howell's novel, a bloody riot, reflects the harsh inequities of capitalism in the late nineteenth century and the class conflict simmering just below the surface of New York society... (Crane 2007, 61).

As Crane argues, realists described their social environment with all the circumstances, changes and conditions they lived in. In that sense, they might not be reliable in respect to a single '*techne*' or '*art*' as Plato would have it, but notice that their writings are nevertheless a valuable source of information on those segments of the culture of their time they write about. We can see them as delivering firsthand witness testimony about the world they lived in. Given that they were primarily concerned with capturing reality in all of its aspects, we have no reason to doubt the veracity of their descriptions. Their interest was almost non-discriminatorily: as Crane argue:

> Realist fiction responded to and participated in a period of sweeping and dramatic transformation following the nation's bloodiest war. Among other things, the post-war era included Reconstruction and its failure, the rise of Jim Crow, unprecedented population growth, revolutions in transportation and communication, a vast influx of immigrants and migration of rural Americans to cities, a turbulent economy characterized by a number of bankruptcies, depressions, and panics as well as by an equally staggering record of economic growth (Crane 2007, 167).

Sceptics might claim that we do not need literary fiction in order to get these kinds of information – as Currie may phrase it, we do not turn to literary fiction to obtain them. We have history books, sociology, anthropology and various other disciplines, all of which are much more reliable than realist literary fiction. Consequently, the epistemic potential of realist novels fades away in comparison to other sources.

The force of this argument is quite plausible: certainly, the practice of literary fiction is not, unlike the practice of scientific discourse, primarily concerned with truthfulness, epistemic reliability, or even the desire or intention on the part of the agents involved to inform. Artistic aims take precedence, both in the case of writers (who pursue artistic aims, such as literary originality or novelty) and in case of readers (who pursue aesthetic pleasure, not a lesson in history). This may very well be the case, but I am not sure that argument along these lines is a decisive argument against epistemic capacities of realist literature. Certainly, if someone is interested in the history of American urbanization, one should look at the history books. However, that in itself does not preclude epistemic dispositions of literature (in this case, transference of factual historical knowledge) once the reader has taken up the book and is engaging with it. The relevant question then becomes, once we engage with *Sister Carrie*, can we take Dreiser to be reliable in what he is saying? Based on the analysis Crane gives us, we can. We then need to explain, in respect *to what* exactly is Dressier reliable? Certainly in describing social circumstances and practices, economical growth, technical innovations, entertainment industry and financial system, to mention but some of the things he writes about. The way to reach these is through factual descriptions, such as this one:

> In 1889. Chicago had the peculiar qualifications of growth which made such adventuresome pilgrimages even on the part of young girls plausible. Its many and growing commercial opportunities gave it widespread fame, which made of it a giant magnet, drawing to itself, from all quarters, the hopeful and the hopeless--those who had their fortune yet to make and those whose fortunes and affairs had reached a disastrous climax elsewhere. It was a city of over 500,000, with the ambition, the daring, the activity of a metropolis of a million. Its streets and houses were already scattered over an area of seventy-five square miles. Its population was not so much thriving upon established commerce as upon the industries which prepared for the arrival of others. (Dreiser 1981, 15).

Given a common reader's general knowledge about urbanization and industrialization, she can take this description as true and learn some facts about Chicago from reading *Sister Carrie*.

So far, in defending the epistemic reliability of literature I relied on the examples from realism, particularly factual descriptions available in these works. I claimed that:

(i) realism, as a literary genre, is to a great extent defined by its norm of truthfully representing the world;

(ii) given that aim, authors are epistemically reliable in what they write. This reliability stems not only from first-hand witnessing the events they were writing about, but also from meticulous researches and observation;

(iii) readers are aware of the conventions governing realistic writings and can therefore accept what they read as truthful

But some problems are obvious. First, how are we to account for epistemic reliability of works outside of realism? Second, does the fact that sometimes factual descriptions within literary works (including realistic novels) contain mistakes diminish cognitivist's claim? Unless these worries are addressed, the case for treating literary works as epistemically unreliable increases. Let's take these in turn.

4. Epistemic reliability and literary genres

How can a literary cognitivist defend epistemic dispositions of works that are not realistic? Part of the answer depends on how we understand realism. When defined chronologically, realism is the dominant paradigm in novel writing in the nineteenth century, including writers like Balzac, Dostoyevsky, Flaubert, Howells, James etc. We saw with Crane that realism can also be characterized in terms of principles and conventions governing the selection of themes (detailed portrayal of reality and psychological workings of the characters) and methods of writing. In realism, unlike in other literary genres, artistic/aesthetic aim is to portray and describe reality: this artistic aim corresponds with the epistemic ideal of reliable informer. In other literary periods and with other genres, these aims do not necessarily collide – in fact, with genres such as romances and futurologies, artistic/aesthetic aims include deliberate rejection or turning away from reality and abandoning epistemic ideal of reliability. To defend the view that literature is cognitively valuable, we have to provide an account of how this turning away from reality can nevertheless have cognitive payoffs.

One way to do so is to focus on genres that, due to their genre-specific themes and literary devices cannot be taken as reflecting reality. What comes to mind is science fiction, a genre characterized by unlimited freedom in imagining things that are not true to reality. Fred Botting talks about science fiction's "unbounded explorations of change, outsiders, escape: its 'freedom of imagery' is freedom from realistic conventions" (Botting 2005, 112). Many other science fiction theoreticians emphasize defamiliarization of the familiar, radical disjunction from the real and cognitive estrangement typical of science fiction. A defining feature of science fiction is that readers have to recognize that what it presents is not real, not actual, not of this world. So how then can science fiction have the epistemic dispositions to reveal to us something epistemically significant about the world?

To answer this, we have to first recognize that, regardless of the literal falsity of descriptions and facts that make up for the fictional world of a science fiction novel, the link with reality is not too distant. As one literary scholar explains, "The basic building-block of science fiction is accordingly the novum – a discrete piece of information recognizable as not-true, but also as not-unlikely-true, not-flatly (in the current state of knowledge)-impossible" (Shippey 2005, 13–4). Shippey's claim is that, for all of its exaggerations, surreal settings, destruction of the world and technological innovations, science fiction deals with deeply important human issues. In doing so, as Seed argues, it "asks fundamental questions about the world and the nature of selfhood" (Seed 2005, 3) that bear importantly on how we conceive of ourselves, other people and the world around us.

To give but a few examples, consider H.G. Wells'repeated return to the topic of genetic engineering, as well as the impact of Darwinism and evolution on people, which raises important questions about the identity and agency of human beings, as well as the boundary between species and the role of science and its impact on humans. Particularly important in this respect are works pertaining to utopias and dystopias, which not only explore different political principles and social arrangements, but also question the nature of human beings, their willingness and abilities to connect with each other and form meaningful relations, not to mention their identity and self-awareness. Philip E. Wegner shows the dependence of utopia on the particular political regime which gives rise to it: "through its presentation of this alternative community, the Utopian narrative has the effect of both highlighting in a negative light many of the problems of the reigning social order, and perhaps even more significantly, of showing that what is taken as natural and eternally fixed by the members of that society is in fact the product of historical development and thus open to change" (Wegner 2005, 80).

Considerations along these lines show that, despite a high degree of the breach from 'reality' and 'here and now', science fiction is deeply rooted in the deepest human concerns and can therefore contribute in substantial ways to our cognitive endeavour of understanding those segments of the world that depend on such concerns. Unlike with realism, the reader of these works has to be more careful in adjusting her level of trust in what she reads. She needs to be careful in navigating the permeable line that separates writers' imagination from how things are, and she needs to focus more specifically on recognizing the underlying themes that these works explore through abandonment of realism. The relevant epistemic gain with these works does not relate to accumulation of factual knowledge; rather, the reader is challenged to reflect on, modify, confirm or abandon her view on particular issues depicted in such works, where such issues are of lasting importance for our understanding of our reality.

5. Literature and the wider intellectual context

We saw that literature presents a valuable source of knowledge of our society, social practices, institutions and human interactions because of how the authors construct fictional worlds so that these reflect the real world. However, it would be a misconception of literature's epistemic dispositions to claim that this is all that authors do. We saw in the last section that epistemic dispositions of literature include also literature's engagement with what Tomas Nagel calls 'the mortal questions'. In this part we will explore how literature does this from the point of view of its relations with other knowledge-seeking disciplines.

A look into our literary tradition is again helpful for this task, and we can begin by exploring literature's link with psychology. Consider Victorian literature. Nicholas Dames (2005) points to a parallel development of psychology as a discipline and a novel as a literary genre, identifying two features they have in common: the overlap of their formal concerns (ways of narrating the self and one's own experience) and the shared goals (gaining a complete image of the mind's processes). Athena Vrettos concurs:

> Indeed, almost anywhere we look in Victorian fiction, we can see the influence of and interest in psychology, ranging from explicit engagements with contemporary philosophies of mind, to theories of character development (...) to fictional experiments with the more speculative branches of Victorian science and pseudoscience – physiognomy, phrenology, and, later in the century, psychical research into telepathy, trances, ancestral memory, and other mysteries of the mind. (Vrettos 2002, 70)

The fact that there is an overlap of thematic concepts and subject matter between literature and psychology is helpful in showing how literary writers can be concerned with themes and issues more frequently associated with scientific domains, but we still need to show that what literary authors say about these concepts has the same epistemic status as what psychologists say – only then can we see the institutional structure of literature as supporting its epistemic aspirations. We have to show that authors were writing from knowledge, rather than from imagination, unbounded by relevant facts. In other words, we have to show that authors had some expertise in what they were saying and that they participated in the same intellectual quest for knowledge as psychologists. In defending such authorial expertise, some critics point to the mutual intellectual environment that writers shared with psychologists, psychologists and other scientists whose influence is visible in their work. These scholars suggest that the interest in psychology as well as profound knowledge that some authors, such as James and Eliot reveal, stem from their friendships and familial connections.[8] Probably the most famous example are brothers Henry and William James. Critics routinely suggest that Henry's interest in and knowledge of psychology can be explained by the fact that he was deeply interested in his brother's field of research and was able to pick up first hand his psychological theories and discuss them with him.[9]

In her analysis of the nineteenth century literature, Janis McLarren Caldwell (2004) explores various ways in which "several influential literary and medical writers were allied in one project, that of negotiating between two distinctly different ways of knowing – between, that is, personal experience and scientific knowledge of the natural world" (2004, 1). Caldwell analyses some of the most influential and well known literary works of the century (*Frankenstein, Sartor Resartus, Wuthering Heights, Middlemarch*) tracing the influences – psychological, philosophical, medical – that shaped the creation of these novels. It is important for us to recognize that with their writings, these authors actively participated in the intellectual debates with scientists and philosophers.[10]

The extent to which literature and philosophy intersect and overlap has lead some philosophers to claim that the two should be merged together and some of the work in contemporary literary aesthetics is dedicated to delineating the lines between the two disciplines.[11] Here I do not want to pursue this debate, but to show the implications of it for the claim that literature is cognitively valuable. Philosophical dimension is indispensable in great many literary works, ranging from 'philosophical romances' of Hawthorne, Poe and Melville, Shakespeare's tragedies, Dos-

8 Here is Vrettos: "Eliot's intellectual circle included both Darwin and Spencer, and she lived with George Henry Lewes, whose p*roblems of Life and Mind* (1879) made important contributions to the field of physiological psychology. Many of Eliot's writings demonstrate the same psychological principles as we find in Lewes's work, and critics such as Gillian Beer and Nancy Paxton have shown how Eliot's fiction offered an ongoing dialogue with the work of Darwin and Spencer" (Vrettos 2002, 71). See also Fessenbecker (2020); Whiteley at all (2024).

9 An exception to this is Meisel (2006), who claims that Henry James in his literary works goes beyond the psychological findings of his brother and includes the domain of unconsciousness. Meisel's analysis of the Modernist literature offers an interesting insight into the way Modern literature relies on psychology, as well as the impact it had on various psychologists.

10 Illuminating example in this context is Caldwell's analysis of Mary Shelley's literary project that gave rise to *Frankenstein*. Though this novel is often seen as the beginning of the science fiction genre, Shelley was in fact inspired by the medical research of her time, committed to exploring the nature of the 'vital principle' which enabled life. This concern did not spring from Shelley's imagination, but from intellectual climate she participated in. See also Whiteley at. all 2024.

11 See Eldridge 2009, Vidmar (2016, 2017, 2019).

toyevsky, Joyce, Kafka, Fitzgerald, science fiction, utopias etc. If literature is primarily conceived of as a humanist discipline, in addressing these questions it joins hands with philosophy.

Psychology, natural science and philosophy are not the only domains that feature in literature: sociological questions are also often inserted in various works. Seeing "the social novel as a form of discourse that can reach into all other areas of social experience" (p. 5), Dominic Head (2002) explains the impact of novel on society in the following way:

> Here there is a direct bridge between seriousness of novels that scrutinize the status quo, and less reflective expressions of popular culture. The post-war novel has done much to discredit a rigid distinction between 'high' and 'low' culture, and, indeed, the prominent protagonists, from Jim Dixon to Bridget Jones – characters that have been rightly seen to typify new social moods – have invariably had popular, or at least middlebrow tastes. The novel, in short, has managed to cultivate a new intellectual space: it is the middlebrow art form par excellence, with unique and unrivalled access to every corner of social life, but a form that retains that 'literary' or serious quality, defined as the ability to deliberate or stimulate reflection on social and cultural questions. (Head 2002, 6)

Head's claim that that novel has "unique and unrivalled access to every corner of social life" explains why we often get the sense of understanding our social reality better when that reality is reflected in works we read. We have to recognize that, to the extent that literature shares thematic concerns with other disciplines, it can contribute to our world-view and the way we come to understand our experience.

6. Mistakes and contradictions in literature

Now that I sketched some directions in which to think of the epistemic dispositions of literature, I have to address two problems that skeptics raise that may diminish our epistemic faith in the novels we read. Both relate to Currie's concerns regarding the lack of proper institutional mechanisms that dictate the epistemic goals of literature. First, does the fact that sometimes literary works contain mistakes diminish the claim of literary cognitivists? Second, how to respond to the problem of contradictions, i.e. the fact that different works offer different, sometimes even contradicting perspectives on the same issue?

Let me start with the problem of errors. While anti-cognitivists attach too much significance to the fact that some factual descriptions are wrong – ignoring the fact that other discourses are not error free and that they too sometimes present incomplete, mistaken and inaccurate perspectives on things – the fact remains that the project of establishing aesthetic cognitivism is diminished by the factual errors which are, much like Currie ad Stolnitz claim, tolerated in literature. Unlike faulty scientific reports which do get corrected, at some point, resulting in dismissal of the originally formulated theory, we generally do not discard works of literary fiction when we spot factual errors in them. This is not to suggest that we do not apply other sorts of mechanisms that signal their epistemic status. One of the most often employed such mechanism is classification of a certain work as fiction rather than as classic, where fiction is less likely to be seen as cognitively valuable and classics are considered important precisely for their treatment of human intellectual concerns. The underlying assumption here is that our encounters with fiction do not invite epistemic engagement; a reader is unlikely to treat descriptions in such works as potentially important for their epistemic aims. This is because she recognizes the triviality of thematic concerns, psychological emptiness of character portrayals and other segments which signal work's disinterest in being cognitively valuable.

Errors however do occur. Consider Philip Larkin's depiction of the breaking of the waves in his poem *Absences*: as critics discovered, Larkin's descriptions is simply wrong – a point that Larkin did not know at the time he composed the poem.[12] In cases like this, author describes a state of the affairs she does not know is factually wrong; therefore, these are false accounts of some segments of our reality. Such distortions can motivate acquisition of false beliefs from the work if the reader does not have knowledge of the relevant facts which shields her from forming false beliefs. Does that make literature generally, and particular works, epistemically infertile?

As a way of response, let me offer several epistemiclly grounded reasons for mistakes in literature. Sometimes mistakes may arise from author's ignorance but often, that ignorance itself is a matter of mistakes in the overall intellectual climate to which the writer belongs. Consider another example that Lamarque discusses: one of the central motives in Henrik Ibsen's *Ghosts* is that of syphilis. As Lamarque claims, "Much of the symbolism in Ibsen's *Ghosts* rests on the moral stigma of syphilis and its being passed from Captain Alving to his son Osvald. But Ibsen and his contemporaries had a quite different conception of syphilis from our own: they were wrong about the heritability of the disease" (Lamarque 2009,

---

12 I take this example from Peter Lamarque, who claims: "There is no artistic intention behind the distortion of fact so no explanation is forthcoming in terms of "artistic license". Here a factual mistake does lead to a literary weakness. It is a failure of mimesis or verisimilitude" (Lamarque 2009).

199). However, if this wrong conception were corrected, the whole moral and symbolic dimension of the play would be lost, to the extent that it would make the play “virtually unintelligible” (Lamarque 2009, 199).

The question now is, does the factual incorrectness of Ibsen’s portrayal of syphilis neutralize the work’s capacity to reveal to us something about the world? I think not – though we do run the risk of forming our beliefs about syphilis from unreliable source. What is important to note is that Ibsen’s error does not reflect his ignorance or prejudice, but the scientific views of the time the work was written – and views that we only now recognize as wrong. Other such examples include mesmerism in Dickens, phrenology in Charlotte Bronte or physiognomy in Trollope. What seems to be a mistake from the perspective of modern readers was not a mistake for the readers contemporary to the author, as our analysis of Victorian novel revealed. While the views expressed in these works are wrong – and therefore can mislead the audience unaware of the relevant scientific discoveries – these works do contain important cognitive dimensions. For one thing, to the extent that literary work embodies scientific, political, religious and sociological beliefs of the time it was written, it can reveal these beliefs to the contemporary audience which might be unfamiliar with them. While such approach to the cognitive value has a kind of ‘archival’ dimension that Gibson points to, notice that often such works help us understand certain practices that were or still are implemented in society. Margaret Mitchell’s depiction of slaves in *Gone with the wind* as irrational and infantile explains some forms of racism that justified (in her days) the institution of slavery. More importantly, mistakes of this kind do not render literary work incapable of being revelatory and illuminating of some aspects of reality, morally complex situations or psychological difficulties. For all the factual errors in representing syphilis, we recognize *Ghosts* as dealing with the problems of dysfunctional families, duties and socially determined roles, and the insights related to these aspects of the novel are not diminished in light of the factual mistakes.

Of course, there are numerous examples of works which seem to signify their reliability but which extend well beyond the regular reader’s knowledge. *My sister’s keeper* abounds in medical information regarding acute leukemia and although its author Jody Picoult repeatedly emphasizes her dedication to research, it is not hard to imagine that some of her claims regarding the medical procedures are not entirely correct. My point however is, the ethical work that the novel is doing – that of inspiring us to reconsider the roles we all have as family members, of challenging us to negotiate our own life plans and projects with that of our family’s needs – are not affected by the work’s potential factual errors. The point is, when discussing the relevance of mistakes for the overall cognitivists' claim, a distinction needs to be made along the lines I sketched: not every mistake is relevant in evaluating a work's epistemic dispositions. The hardest cases for literary cognitivist are those in which the author deliberately deviates from facts without making it obvious. Particularly in domain of ethics, such errors can have rather damaging consequences. Should a reader, upon engaging with *Gone with the wind,* come to believe that black people were well off prior to the war and abolitionism, we would have a good reason to consider the book misleading.

Similar reasoning applies to the fact that different works present contradictions. Sometimes, contradictions are the result of works reflecting scientific (i.e. political, religious, sociological, philosophical etc) attitudes and beliefs accepted (presumably) by the author and his intended audience. In some cases, writers intentionally contradict the established world view of the community in order to challenge the accepted morality and overall ideology. The fact that literature generally contains contradictions which are not 'corrected' can also be explained by the fact that literature not only reflects, but also, and more often, criticizes and challenges not only scientific changes but general attitudes of the public towards specific issues. Philosophical, social, political, scientific and religious attitudes change. Literature reflects that change and in a very important way, serves as an archive of the general growth and development of human body of knowledge.

7. Conclusion

My aim in this paper was to defend the view that literature is cognitively valuable and I did so against one particular challenge issued by Currie: the one undermining epistemic reliability of literature. My approach was based on extensive analysis of critical commentaries on various literary works, their writers, and the overall literary tradition, all of which are concerned with showing ways in which literary fiction can contribute to our cognition. Arguably however, none of them would diminish Currie's skepticism. Indeed, if we were to argue that the epistemic dispositions of literature are the same as those of science, we would be wrong. But on my view, that does not diminish our right to treat literature as epistemically potent.

My argument was for the most part based on pointing to the aesthetic and epistemic norms underlying particular periods and genres in our literary tradition. I argued that, to the extent that the readers are aware of at least some of the conventions guiding literary production in certain periods, they can rely on some set of epistemic criteria, however loosely defined, which can regulate the level of trust they give to a particular work. It is not, I think, an exaggeration to say that this knowledge is available to readers, particularly those who repeatedly turn to literature because they enjoy the intellectual stimuli provided by it, and the aesthetic rewards they receive in response. As Eva-Maria Konrad argues, "there are a number of rules and conventions that we have to learn if we want to either read or write a fictional text in the correct way. These conventions are not established by a single readers or authors alone, but by the whole literary community" (2017, 47).

My account also draws from the rich tradition of literary criticism. For reasons of space I could not provide a substantial account of the epistemic structure of literary criticism, but the examples offered can at least mitigate some of Currie's claims. His claim that institutions of criticism are not "designed to promote, nor are they connected in any systematic way with, epistemic reliability" (2020, 185) seems plausible, but we should also consider the efforts that literary critics and scholars employ in engaging with individual works in order to trace their epistemic source: obviously, only a critic with sufficient expertise in the relevant domains can offer such insightful analyses. In a similar way, Currie's generalized claim that "when truthfulness is invoked in criticism, it often seems to mark the achievement of some vivid effect, with no very obvious connection to the idea of being right about anything in particular" (2020, 185) can easily be met with counter-examples. Recall Gibson's meticulous depiction of the insight provided by Faulkner regarding the racial issue, Crane's analysis of realists' interest in the everyday social world, or Vrettos account of the psychological domains explored by literature. On my view, such examples show that literary critics make a good job in underlying the particular epistemic segments of different works and are good in pointing to their reliability.

Of course, as several anti-cognitivists are quick to claim, readers, for the most part, rarely check such 'extra-literary' writings and they commonly engage with works relying on their own sense of 'what the works tells them' rather than on what critics find reliable. On Currie's view, this is additional reason to doubt literature's cognitive capacity, which eventually leads him to suggest that perhaps readers only pretend to have learnt something. I do not think that is right however. I have elsewhere (2024) proposed to think of our learning from literature as accommodated within 'peer-review' model: the one which allows individual reader a certain liberty in evaluating what segments of a work, if any, offer potential for learning. For some readers, this may amount to few facts, for others, the moral or psychological principles depicted in a work may be intellectually rewarding, yet for some others, the particular way in which a particular character responds to a moral dilemma or existential crisis. My intention is not to diminish author's reliability or expertise, only to provide a model of learning from fiction more aligned with the basic point that Currie makes; the one regarding the subjective aspect of literary experiences. What one picks up from literature depends more on the reader than on the particular work: some dispositions may go unnoticed, while some may be activated and taken in directions not predicted by the writer. While on Currie's view such a subjective nature of our responses to literature is yet another reason to doubt its epistemic value, on my view it is precisely the aspect that makes it so immensely important, helpful and significant in our individual attempts to make sense of ourselves, our experiences and other people.[13]

Bibliography

Appleman, Deborah. *Literature and the New Cultural Wars*. 2021.

Aristotle, *Poetics*. In *The Complete works of Aristotle*, ed. By Jonathan Barnes, vol. 2. 1984.

Botting, Fred. "'Monsters of the Imagination': Gothic, Science, Fiction." In *A Companion to Science Fiction*, edited by David Seed, 112–126. Blackwell Publishing, 2005.

Carroll, Noël. "Literary Realism, Recognition, and the Communication of Knowledge." In *A Sense of the World: Essays on Fiction, Narrative and Knowledge*, edited by John Gibson, Wolfgang Huemer, and Luca Pocci, 24–41. Routledge, 2007.

Crane, Gregg. *The Cambridge Introduction to the Nineteenth-Century American Novel*. Cambridge University Press, 2007.

Currie, Gregory. *Imagining and Knowing: The Shape of Fiction*. Oxford University Press, 2000.

Dames, Nicholas. "The Withering of the Individual: Psychology in the Victorian Novel" in A Concise Companion to the Victorian Novel, ed. O'Gorman F., MA: Blackwell Publication, pp. 91–112, 2005.

Davies, David. "Thought Experiments and Fictional Narratives." *Croatian Journal of Philosophy* 7, no. 9 2007: 29–45.

13 This work has been supported by the Croatian Science Foundation under the project number UIP-2020–02–1309. The previous version was presented at the conference Balkan Analytic Forum, held in Belgrade in 2024. I am grateful to the audience for their insightful comments, and to prof. Miroslava Trajkovski for her generous support of my work.

Diamond, Cora. "Henry James, Moral Philosophers, Moralism." In *A Companion to the Philosophy of Literature*, edited by Gerald L. Hagberg and Walter Jost, 268–284. Wiley-Blackwell, 2010.

Dilman, Ilham. "Dostoyevsky: Psychology and the Novelist." In *Philosophy and Literature*, edited by A. P. Griffiths, 95–114. Cambridge University Press, 1984.

Eldridge, Richard, ed. *The Oxford Handbook of Philosophy and Literature*. Oxford University Press, 2009.

Fessenbecker, Patrick. *Reading Ideas in Victorian Literature: Literary Content as Artistic Experience.* Edinburg University Press, 2020.

Gibson, John. *Fiction and the Weave of Life*. Oxford University Press, 2007.

Gibson, John. "Literature and Knowledge." In *The Oxford Handbook of Philosophy and Literature*, edited by Richard Eldridge, 466–485. Oxford University Press, 2009.

Gibson, John, ed. *Narrative, Emotion and Insight*. University Park: Pennsylvania State University Press, 2011.

Goldman, Alan. *Reasons from Within: Desires and Values*. Oxford University Press, 2009.

Hamalainen, Nora. "Literature and Moral Change: Rapture, Universality of Self-Understanding." In *Narrative and Self-Understanding*, edited by G. Hagberg, 29–51. Palgrave, 2019.

Head, Dominic. *The Cambridge Introduction to Modern British Fiction*. Cambridge University Press, 2002.

Kieran, Matthew. "Art, Morality and Ethics: On the (Im)Moral Character of Art Works and Inter-Relations to Artistic Value." *Philosophy Compass* 1, no. 3 (2006): 191–200.

Kjeldgaard-Christiansen, J. "Do Dark Personalities Prefer Dark Characters? A Personality Psychological Approach to Positive Engagement with Fictional Villainy." *Poetics* 85 (2021).

Konrad, Eva-Maria. "Signposts of Factuality: On Genuine Assertions in Fictional Literature." In *Art and Belief*, edited by Sullivan et al., Oxford University Press, 2017.

Ladenson, Elizabeth. *Dirt for Art's Sake: Books on Trial from Madame Bovary to Lolita*. Cornell University Press, 2007.

Lamarque, Peter. *The Philosophy of Literature*. Blackwell Publishing, 2009.

McGregor, Rafe. *Narrative Justice*. Rowman & Littlefield, 2018.

McLaren Caldwell, J. *Literature and Medicine in Nineteenth-Century Britain: From Mary Shelley to George Eliot*. Cambridge University Press, 2004.

Meisel, Perry. "Psychology." In *A Companion to Modernist Literature and Culture*, edited by David Bradshaw and Kevin J. H. Dettmar, 79–91. Blackwell Publishing, 2006.

Nussbaum, Martha. *The Fragility of Goodness*. Cambridge University Press, 1986.

Plato (1997), *The Republic*, in *Plato, Complete Works*, ed. Cooper J. M., Hackett Publishing Company, pp.971–1223Seed, David, ed. *A Companion to Science Fiction*. Blackwell Publishing, 2005.

Shippey, Tom. "Hard Reading: The Challenges of Science Fiction." In *A Companion to Science Fiction*, edited by David Seed, 11–26. Blackwell Publishing, 2005.

Stock, Kathleen. "Fiction, Testimony, Belief, and History." In *Art and Belief*, edited by Sullivan et al., Oxford University Press, 2017.

Stolnitz, Jerome. "On the Cognitive Triviality of Art." *British Journal of Aesthetics* 32, no. 3 (1992): 191–200.

Vidmar, Iris. "Challenges of Philosophical Art." *Proceedings of the European Society for Aesthetics* 8 (2016): 545–569.

Vidmar, Iris. "On Literary Cognitivism from the Perspective of Difference Between Literature and Philosophy." *Synthesis Philosophica* 64, no. 2 (2017): 371–386.

Vidmar, Iris. "Rethinking the Philosophy-Literature Distinction." *Rivista di Estetica* 70/1 (2019): 156–171.

Vidmar Jovanović, Iris, ed. *Narrative Art, Knowledge, and Ethics*. Rijeka: Filozofski Fakultet, 2019.

Vidmar Jovanović, Iris. "Unsuitability of Literature for Moral Education." *Theoria* 90 (2024): 417–428.

Vrettos, Athena. "Victorian Psychology." In *A Companion to the Victorian Novel*, edited by Patrick Brantlinger and William B. Thesing, 67–83. Blackwell Publishing, 2002.

Wegner, Phillip. "Utopia." In *A Companion to Science Fiction*, edited by David Seed, 79–94. Blackwell Publishing, 2005.

Whiteley, Giles, Monika Class, Cian Duffy, Andrea Selleri, and Peter Garratt, eds. *Literature and Philosophy in Nineteenth-Century British Culture*, Routledge, 2024.

Winner, Ellen. *How Art Works: A Psychological Exploration*. Oxford University Press, 2018.

Young, James. *Art and Knowledge*. Routledge, 2001.

Young, James. "Literary Fiction and True Beliefs." In *Art and Belief*, edited by Sullivan et al., Oxford University Press, 2017.

https://doi.org/10.18485/baf.2025.2.ch6

Ana Kuburić Zotova

HOW MINDFULNESS AS A DISPOSITION IMPACTS EPISTEMIC DISPOSITIONS

Abstract: This paper explores how a mindfulness disposition (defined as inclination and inner capacity of paying sustained, non-reactive attention to present-moment experience) influences cognitive information processing and epistemic performance. Drawing on a systematic review of psychological, neuroscientific, psychotherapeutic, and philosophical literature, it examines the cognitive effects of mindful awareness of experience. The reviewed studies report improvements in perceptual discrimination and comprehensiveness, interpretative accuracy, source evaluation, cognitive flexibility, and openness to alternative hypotheses. These processes are foundational and prerequisite for the development of key epistemic dispositions discussed in virtue epistemology and critical thinking theory – a topic that has rarely been addressed. The paper argues that while a mindfulness disposition enhances epistemic competence, its benefits become especially pronounced when combined with analytical and proactive skills. The analysis also considers parallels between secular mindfulness and experiential approaches in phenomenology, hermeneutics, and pragmatism, suggesting broader interpretative relevance.

Key words: epistemic dispositions, mindfulness, present-moment experience, cognitive flexibility, open-mindedness, perception, awareness.

Introduction

All people, at different stages of life, struggle to varying degrees with emotional regulation, interpersonal relationships, and cognitive overactivity. Mindfulness is increasingly recognized as a potential method for enhancing both reasoning capacities and emotional regulation. Numerous contemporary psychological studies on the effects of training non-reactive attention to the flow of present-moment experience suggest that it facilitates the deautomatization of emotional, behavioral, and cognitive patterns that are not functional, thereby enabling more appropriate responses to situations.

While individual aspects such as attention, emotional regulation, meta-awareness, and deautomatization have been studied in the context of mindfulness, their epistemological implications have been less frequently explored—this is the aim of the present paper. A review of psychological and psychotherapeutic literature provides strong indications that mindfulness can support key intellectual dispositions, such as: reliable perception, cognitive flexibility, the disposition to consider situation and context, the ability to question one's own beliefs, to recognize one's own standpoints, the courage to explore seemingly complex or unpleasant ideas, and openness to others' points of view. The hypothesis from which this work begins is that the training of mindful awareness—defined here as non-reactive attention to the unfolding of present-moment experience—enhances a set of dispositions that are prerequisites for successful epistemic reasoning. If this is the case, then mindful awareness, when integrated with analytical and proactive thinking, may foster more successful epistemic reasoning. More precisely, this paper aims to explore how mindfulness training affects four interrelated domains of cognitive and emotional functioning: perceptual openness, meta-awareness, cognitive flexibility, and emotional regulation, and to consider these effects in the context of developing epistemic dispositions—an area in which such connections have rarely been examined.

The paper begins by defining mindfulness as a disposition and highlighting its conceptual parallels with phenomenological, hermeneutic, and pragmatic approaches to experience, as well as with the experiential focus in Gestalt and client-centered therapies. It then reviews psychological and neuroscientific findings on the impact of mindfulness on attention, meta-awareness, deautomatization, and emotional regulation, analyzing how these findings relate to epistemic dispositions essential for epistemic success in the frameworks of virtue epistemology and critical thinking theory. Finally, the paper addresses the limitations of the mindful state and proposes its integration with analytical and proactive skills.

1. Defining Mindfulness Disposition

Mindfulness is a distinct form of awareness that we can engage in at certain times. This form of awareness is characterized by:

(1) Directing attention to present-moment experiences. This is how one enters a mindful state. John Kabat-Zinn's operational definition says: "Mindfulness is the awareness that emerges through

paying attention on purpose, in the present moment, and non-judgmentally to the unfolding of experience moment by moment" (2003, p. 145).

(2) This is not just any kind of attention – but a sustained, non-reactive attention as described by Evan Thompson, often referred to as non-judgmental attention according to Kabat-Zinn.

I will quote Thompson's explanation, which states that "non-reactive" means that "ideally one simply observes or witnesses the coming and going of sensory and mental events without getting caught up in cognitive and emotional reactions to them" (2015, pp. 52–53).

In other words – everything that enters the perceptual field is observed and then released. If something triggers an emotion, a mental comment, or act, it is noticed and let go. In this way, a person can see, hear, feel things they are not in the habit of noticing or may not want to notice. This ability, which children are born with and use to learn about the world around them, can weaken over time. Fullness of awareness is characterized by the breadth of attention and the ability to willingly direct attention, deciding where to focus and for how long.

(3) In this way, everything that can be part of the phenomenal field of present-moment experience is noticed: sensory information from the environment or signals from the body, as well as mental content (emotions, thoughts of past, now and future), whether they are bad, good, or neutral. However, all of these are released – the person does not get attached to any particular contents but surrenders to the flow of consciousness, allowing the unfolding of experience moment by moment.

(4) Through mindful awareness, new content is perceived, new distinctions discerned, new connections among phenomena recognized, new categories created, and old ones can be recategorized (Langer, 1989, pp. 63–67).

The path from state mindfulness to trait mindfulness and disposition will not be my focus here.[1] Instead, I will begin by examining how an already developed mindfulness disposition can positively influence certain epistemic dispositions. By mindfulness disposition I will consider inclination and inner capacity of paying attention to present-moment experiences with a non-reactive attitude.

1 Readers may explore Kiken et al. (2015) and Sternberg (2000) for aforementioned context.

2. The Similarity of Paying Attention to the Stream of Experience Across Different Philosophical, Psychological Traditions and Mindfulness Meditation

Based on my earlier research on the use of the term 'experience' (Kuburic, 2022), I found that the founders of all three philosophical traditions—hermeneutics, phenomenology, and pragmatism—emphasized the importance of paying attention to the stream of experience. Gadamer and Dilthey believed that paying attention to the stream of experience enhances understanding and contributes to epistemic success. Husserl advocated the idea that a careful examination of the stream of experience *(Erlebnisstrom)* helps in the formation of meaning. In *Principles of Psychology*, James considers how human interests and selective attention interact with the stream of experience.

I believe that the distinctive viewpoints from which different philosophical traditions observe the stream of experience complement each other in clarifying the structure and functioning of experiencing. They explain the relationship between various cognitive and psychological processes and experience itself. Although their methodologies differ, as do their expectations regarding the increased attention to the process of experiencing, the insights drawn from their work about the structure and functioning of experiencing is mutually consistent and align with Gestalt psychotherapy, client-centered therapy, focus-oriented psychotherapy and Antonio Damsio's neuroscientific standpoint.

Mindfulness can be added to these traditions that nurture a relationship with one's own process of experiencing, though in different ways. Thompson and co-authors in *The Embodied Mind* derive the concept of experience from the mindfulness tradition (2016, p. 236), and in later works, also from phenomenology. My intention is to assert more about the way certain epistemic dispositions develop through the mindfulness approach, which I have chosen due to the extensive body of psychological and neuroscientific studies on its effects. However, due to the similar emphasis on paying attention to the stream of experience by the mentioned philosophical and psychotherapeutic approaches, the findings regarding the influence of mindfulness disposition on epistemic success may also be interpreted more broadly. Attaining a mindful state is something natural to us because it concerns, above all, managing non-reactive attention, which is a natural human capacity. This state is not something exclusively achieved through formal meditation, although it is probably most effectively developed through it (Kabat-Zinn, 1994, p. xvii). People vary in their level of mindfulness regardless of culture, environment (rural or urban), or historical period, even though some contexts may be more favorable for its cultivation.

3. Advocates of mindfulness argue

Advocates of mindfulness argue that a changed attitude towards the process of experiencing in the present moment enhances emotional regulation, the ability to control attention, cognitive flexibility, creativity, memory, and the regulation of conation components. In this text, I will focus on cognitive functions enhanced by mindfulness training, explaining them in the context of epistemic abilities and dispositions.

A meta-analysis by Sedlmeier and co-authors analyzes 163 psychological studies conducted on non-clinical population from 1970 to 2011. By calculating the effect size (Cohen's d), they found strong evidence that meditation practice is associated with a range of cognitive and emotional benefits, often showing medium to large effect sizes. Effect size tells us whether the impact can be considered relevant in a practical sense. According to Cohen, a medium effect is visible to the naked eye of a careful observer (Cohen, 1992, p. 156). The effect size for mindfulness, perception, attention and cognition, are shown to be medium (0.28). Meditation has a greater impact on interpersonal relationships (0.44) and the regulation of negative emotions (0.34) (Sedlmeier et al, 2012, pp. 17–18), which is indirectly relevant for the development of epistemic abilities such as perceptual comprehensiveness and openness to alternative hypothesis.

The mindfulness variable included 3,074 participants from 26 studies; the perception variable included 953 participants from 15 studies; attention, 1,307 participants from 22 studies; and cognition, 470 participants from 7 studies. Although most original studies had small samples, the meta-analysis aggregates them into large sample sizes and thus great accuracy of the findings.

4. Epistemic Dispositions Impacted by Mindfulness Training

I will aim to connect the abilities measured by psychological tests – such as attention, meta-awareness, deautomatization, and emotional regulation – with terminology familiar to us when discussing epistemic dispositions. Turri (2015) notes that, within contemporary epistemology, there is broad agreement that knowledge must proceed from reliable processes, abilities or dispositions (p. 529.) On this point, he interprets Ernest Sosa as arguing that knowledge must be produced by a competence – a disposition that would ensure (or make highly likely) the success of any relevant performance under appropriately normal conditions (2007, p. 29). Competence is a disposition to succeed, but not every disposition to succeed is a competence. (Sosa, 2017, pp. 193,195.) "Disposition can be a 'competence' only if its contained source is sufficiently reliable, at least in its distinctively appropriate conditions." (Sosa, 2007, p. 106.)

4.1. Quality of Perception Improved by Training of Attentional Functions

The connection between epistemology and mindfulness is primarily evident in how we grasp information and acquire knowledge based on our experiences, whether those experiences are gained through our perception of physical reality, interaction with others, or exposure to artistic and intellectual works. In all these cases, we rely on the deliverances of our senses, memory, and inference, which create a moment-by-moment experience. Here, I refer to Sosa's notion of deliverances: "Traditionally our knowledge is said to have 'sources' such as perceptions, memory, and inference. Epistemic sources issue 'deliverances' that we may or may not accept."[2] (Sosa, 2007, p. 101) "Acceptance of a deliverance [...] constitutes knowledge only if the source is reliable, and operates in its appropriate conditions, so that the deliverance is safe, while the correctness of one's acceptance is attributable to one's epistemic competence." (Sosa, 2007, p. 103)

In this subsection, I will focus on perception as an epistemic source of information, – specifically, on how mindfulness meditation enhances our ability to register and evaluate perceptual deliverances. According to Sosa, perception belongs to a type of epistemic competence that consists of "dispositions to host a specific range of deliverances under certain coordinated conditions" (2007, p. 106). Another type of epistemic competence related to perception is the "disposition to accept such deliverances at face value, absent any sign to the contrary", in other words, "to implicitly trust the source" (p. 106).

Perception is enhanced through meditation because it exercises various attentional functions. Mindfulness meditation, along with other secu-

2 Sosa at this point only talks about delivering of a certain propositional content and "such deliverance is safe *outright* provided it would then so deliver its content only if true. A deliverance is safe *dependently* on some further fact if, and only if, though not safe outright, it would still so deliver its content, *in the presence of that further fact*, only if true." (Sosa, 2007, p. 101)

larized forms of mindfulness training, can indeed be viewed as "attentional regulatory training regimes," as pointed out by Lutz and colleagues (2008, p. 163). All functions of attention are practiced, but I will highlight three pointing to the path of improving perception: *sustaining, inhibiting*, and *allocating attention*. By training these functions in a non-reactive manner, perceptual abilities are improved, specifically in terms of perceptual discrimination, accuracy, comprehensiveness, and the ability to evaluate situations, contexts, and the reliability of sources.

If we simplify the mechanisms[3], we can say that practicing sustained attention and ignoring non-target stimuli results in higher perceptual accuracy over time (Palmer, 1999, pp. 538–541). For example, meditators give a higher number of correct answers on the N-back task[4] than control groups (Fabio & Towey, 2017, p. 83; Zeidan et al., 2010, p. 603).[5]

Second, better performance in allocating attention during moment-to-moment experiences leads to improved perceptual discrimination over time.[6] For example, meditators show better attention allocation in At-

3 We will simplify the discussion by focusing on a key factor contributing to the development of perceptual accuracy, even though it operates in interaction with other cognitive functions. The cited studies provide insights into additional factors (e.g. working memory, Zeidan et al., 2010, p. 603). Likewise, we do not explore the interrelation between different attentional processes and their joint contribution to perceptual accuracy (e.g., *attention switching* returns wandering attention to the target object, and *attentional inhibition* suppresses less relevant stimuli that emerge; Fabio & Toway, 2017, p. 82).

4 In the N-back task, participants are shown a series of stimuli one at a time. For each stimulus, they must determine whether it matches the one presented **N** steps earlier in the sequence. The value of **N** can vary, requiring them to compare the current stimulus with one presented 1, 2, 3, or more trials before. Demo of test: https://www.psytoolkit.org/experiment-library/touch_nback2.html (accessed January 25, 2025).

5 I consider it justified to refer to studies involving participants trained in focused attention meditation (FAM) techniques—such as the study by Fabio and Towey—when discussing the effects of different attentional functions on perception. This is because mindfulness mediation (MM) practice typically begins with focused attention (FA) training, which later evolves into open monitoring. The capacity to inhibit distractions and to develop atentional monitoring is necessary for transitioning into the 'open monitoring' practice. Without cultivating these capacities, one cannot engage in non-reactive awareness of the stream of experience. As to Lutz et al. (2008) note, MM encompasses both FA and OM practices, where FA involves sustaining attention on a chosen object, and OM entails non-reactive monitoring of the content of experience from moment to moment. Therefore, findings from FAM practitioners are relevant for understanding the foundational attentional mechanisms cultivated in MM as a whole.

6 As with sustained attention, we set aside the interrelation between different attentional processes that, in synergy, contribute to perceptual discrimination (e.g., sustained attention and the accurate working memory representation of the nontarget support the perceptual identification of the target during discrimination tasks; see MacLean et al., 2010, p. 837).

tentional Blink tasks[7] and higher visual search abilities in Visual search tasks[8] (Fabio & Towey, 2017, pp. 79, 80, 83; van Leeuwen et al., 2009, pp. 598). The referenced studies employing N-back, Attentional Blink, and Visual Search tasks primarily investigate visual perception, although other studies have found that meditation-based attention training also affects other types of perception.

Third, practicing attention shifting leads to improved perceptual comprehensiveness, as it involves the ability to move between different aspects of a scene, which can help in recognizing various contexts, connections and meanings. Fabio and Towey (2009, p. 81) found significantly higher self-assessments among long-term meditation practitioners on the *observation* subscale, which assesses the degree of careful attention to a range of stimuli, including exteroceptive, proprioceptive and interoceptive sensations. Using the Kentucky Inventory of Mindfulness Skills, a higher level of observation skills was identified, in the sense that participants reported seeing, hearing, and feeling more.[9] One reason why the effect size of mindfulness on interpersonal relationships is so large is that mindful individuals notice more about others' states, as well as the state of the environment, the general context, and the situation. While this paper focuses on a subset of tasks and self-assessment scales, a variety of other tests are available to assess attentional and perceptual processes, and their results show similar findings.

Robert H. Ennis presents the consideration of the situation as one of the six basic elements in the process of judging arguments (1996, p. 7), and as a disposition essential to critical thinking. As such, it belongs to the group of dispositions that help individuals to represent both their own and others' positions honestly (1996, p. 9). Every belief or decision is part of a broader situation. According to Ennis, the situation shapes the

---

7 In the presence of a sequence of visual stimuli presented in rapid succession at the same spatial location, a subject can typically identify the first target without difficulty, but has difficulty in identifying the second, if it is presented between 200 and 500 ms after the first. Attentional Blink refers to this brief period of reduced processing capacity. See demo of test: https://www.psytoolkit.org/experiment-library/experiment_ab.html (accessed January 25, 2025).

8 "The VS task requires participants to determine the presence of a target (e.g., a letter, a shape or an image) among a number of distracting stimuli." (Fabio & Toway, 2009, p. 76)

9 *The KIMS* was introduced by Baer et al. (2004). The Observe subscale includes items such as: "I pay attention to sounds, such as clocks ticking, birds chirping, or cars passing"; "I notice visual elements in art or nature, such as colors, shapes, textures, or patterns of light and shadow"; "I notice when my moods begin to change"; "I pay attention to how my emotions affect my thoughts and behavior"; "When I'm walking, I deliberately notice the sensations of my body moving".

significance of the thinking activity, some rules that guide it, the meaning of what the thinker is judging or doing, and reveals important factors that should be taken into account. The situation may include the people involved, the physical environment, and the social environment. (Ennis, 1996, p. 7) Ennis's situational considerations imply reflective deliberation that differs from – and builds upon – the mindful grasping of the situation under discussion here. When Shaun Gallagher, drawing on John Dewey's notion of situation, explains the distinction between environment and situation, he emphasizes that reflective attempts to grasp the situation modify it by introducing a reflective element into it. (2020, p. 13). Evan Thompson notes a similar effect of reflection on experience, asserting that there is an unreflected experience prior to reflection (2007, p. 464). The mindful meta-awareness engaged during the mindfulness state of awareness does not allow for the introduction of reflective deliberation into the grasping of the situation. The central aim of the present paper is to explore how non-reactive attention to the unfolding of present moment experience affects emergent reflective activity – activity that implies full-capacity higher-order cognitive processing, which is crucial for logical-analytical evaluation of arguments. As the cited studies indicate, mindfulness fosters set of dispositions that are prerequisites for successful epistemic reasoning. In mindfulness training, complex reflective activities are temporarily deferred to enable clearer perception of the situation. I will go on to explain why successful epistemic reasoning requires the cultivation of experiential consciousness, meta-awareness monitoring in a mindful manner, and the deferral of both reflective deliberation and emotional reactions.

4.2. Questioning Beliefs and Reasons and Recognizing Standpoints Through Meta-Awareness

In the previous subsection, we saw that by enhancing attention and perception, mindfulness meditation helps a person have a more vivid and integrated experience (*experiential consciousness*)[10] of the events in which they are involved. This subsection addresses how mindfulness meditation increases the quality of one's explicit awareness of the contents of experience (*meta-awareness*). We can live through experience without realizing it. Meta-awareness is "the process of directing attention toward the contents of consciousness, thereby gaining an appraisal of the contents of consciousness." (Chin & Schooler, 2009, p. 33) Many authors define meta-awareness as a reflective access to ongoing mental states and processes

10 Following Schooler and Mauss (2010, p. 244), I use the term *experiential consciousness* to refer to "the contents of ongoing experience", as distinct from meta-awareness.

(e.g. Zeidan et al, 2010, p. 603; Smallwood, 2007, p. 527), rather than in terms of attention, as it is framed in the cited definition. In this way, priority is given to understanding meta-awareness as a self-referential relation of consciousness to its own content in real time. But such a formulation calls for greater precision, especially in the context of mindfulness. Although the concepts of meta-awareness and reflection are related, they differ in important ways and should not be conflated. First, reflection may accompany ongoing experience, as in a form of meta-awareness, but it also often involves a disengagement from the flow of present-moment experience, unfolding retrospectively.[11] Even more important, they can interfere because reflection can extinguish or overshadow the flow of experience that mindful meta-awareness aims to register. Second, within the mindfulness tradition,[12] meta-awareness is described as a "non-elaborative, non-judgmental noticing" or witnessing of the contents of consciousness by awareness, without insight in the sense of deliberation. Mindful meta-awareness insights do not originate from discursive thinking (e.g. deliberation, interpretation, analysis), as reflective insights do, but rather from present awareness.

Compared to reflection, mindful meta-awareness involves less labeling, less categorization, less conceptual structuring, less evaluation, and less analysis. Instead, it relies on noticing the arising and passing of mental and physical states and processes, as well as on noticing of their order of appearance and relational patterns. This distinction is important, as not everything that appears in consciousness—whether perceptions, thoughts, emotions, or volitional impulses—calls for commentary. A kind of selection takes place here, one that is often weakened by constant verbalization and cognitive hyperactivity. As Zeidan et al. (2010, p. 598) note, based on Posner and Rothbart (1998), "improvements in meta-cognition are related to the ability to restrict bottom-up processing of exogenously/endogenously driven, task-irrelevant information." Or, in other words, the ability "to 'release' cognitive appraisals of irrelevant information" (p. 598).

Cognitive hyperactivity can take many forms - such as rumination, hypermentalization, mental proliferation, excessive mental elaboration,

---

11 For further reading on two types of dissociations between experiential consciousness and meta-consciousness—temporal and translational dissociations—see Chin & Schooler (2009).

12 It should be noted that there are different types of meta-awareness (for details, see Chin & Schooler, 2009), and here we are referring specifically to mindful meta-awareness. Dunne, Thompson and Schooler in their 2019 joint paper, propose the notion of "sustained (not intermittent) and non-propositional (not involving an internally verbalized judgment) meta-awareness" as a construct that helps clarify the type of meta-awareness cultivated within mindful meditation (p. 308.).

and hyper-reflexivity. The first four represent maladaptive thinking patterns serving distinct psychological functions, while the latter constitutes maladaptive metacognition and introspection.[13] Certain defence mechanisms, such as intellectualization and rationalization, also manifest as cognitive overactivity. All these forms of cognitive hyperactivity share the following features: they are compulsive and patterned; meta-cognition is active only in registering the emergence of thoughts but not their dissolution (resulting in continuous chaining); and meta-awareness does not register the intersections and temporal sequences of thoughts, emotions, desires, and bodily sensations. Studies on the effects of various psychotherapeutic approaches suggest that their common remedy lies in therapies that strengthen attention to the contents of consciousness in real time and integrate cognitive, emotional, conative and bodily states and processes – such as mindfulness-based approaches and other phenomenological therapies (e.g. Gestalt, embodied and existential approaches). Individuals (in our context, scientist and philosophers) who cultivate non-reactive, sustained attention to present-moment experience develop the ability to attend and monitor the unfolding of diverse cognitive, emotional, and conative contents. In doing so, they exhibit not only meta-cognition but a broader meta-awareness that includes meta-cognitive awareness, but also meta-emotional, meta-sensory, meta-conative, meta-behavioural and meta-bodily awareness. During the monitoring of the arising and passing of mental events, selection occurs by releasing irrelevant emotional, conative, and cognitive appraisals. This releasing does not necessarily involve reflective insight or deliberate rejection of content, but rather a non-conceptual letting go that occurs within mindful awareness. Grabovac et al. compare the mechanism through which CBT interventions reduce mental proliferation with that of mindfulness: “Metacognition involves focusing

13 Rumination is a repetitive thinking pattern in which the same emotionally charged content is re-experienced. Each new thought intensifies the primary emotion and can easily escalate a minor concern into a catastrophic scenario. This is commonly observed in depression and anxiety.

Hypermentalization is a mode of thinking characterized by the excessive attribution of mental states to others, despite limited evidence. It functions as a strategy to reduce uncertainty in social relationships and ambiguous interpersonal situations.

Mental proliferation refers to the uncontrolled unfolding of associative thinking, in which a present experience rapidly triggers memories, future projections, labeling, and interpretative elaboration. This cascade of loosely connected thoughts gives a sense of insight and serves the purpose of processing experience.

Excessive mental elaboration refers to the tendency to construct multiple interpretative layers or engage in scenario-building, often extending far beyond situational demands. While this may create an illusion of sophisticated understanding, it frequently functions to avoid the discomfort of uncertainty or emotional vulnerability.

attention on a stream of mental events (mental proliferation) and volitionally interrupting that stream with a new series of mental events whose objects are the preceding thoughts that were part of the mental proliferation. [...]. In contrast, insight and its side effects [such as equanimity] are non-conceptual and non-cognitive in their origin and result in reductions in attachment/aversion or mental proliferation without requiring any cognitive intervention or processing." (2011, pp. 161–162). CBT intervenes by replacing less skillful thoughts with more adaptive ones through metacognitive awareness. In contrast, mindful insights operate at a more fundamental level and lead to healing without cognitive reprocessing.

Grabovac and co-authors propose the BPM model (Buddhist Psychological Model) of the mechanism of change through mindfulness, in which three (I would add, mindful meta-awareness) insights are key – namely, that "all sensory impressions and mental events" have three characteristics: "1. Sense impressions and mental events are transient (they arise and pass away); 2. Habitual reactions (i.e., attachment and aversion) to the feelings of a sense impression or mental event,[14] and a lack of awareness of this process, lead to suffering; 3. Sense impressions and mental events do not contain or constitute any lasting, separate entity that could be called a self."[15] (2011, p. 156). According to their view, the realizing of these three characteristics leads to a reduction of symptoms (e.g. mental proliferation, emotional agitation, strong desire/willing), and an improvement in well-being.

Mindful meta-awareness plays a role in the foundational stages of recognizing beliefs and becoming aware of one's own standpoints. Robert Ennis refers to the disposition to be "reflectively aware of their own basic beliefs" (1996, p. 9). This disposition, along with the mentioned consideration of the situation, belongs to the group of dispositions that help individuals to represent both their own and others' positions honestly (1996, p. 9). We are faced with a complex requirement, because it is only when we become aware of our beliefs – which are often implicit, unconscious, or automatic – that we can begin to question them. Mindful meta-awareness enables the noticing of automatic thoughts that reflect core beliefs, as well as intermediate beliefs (such as rules, assumptions, and accompany-

---

14 Attachment and averstion emerge in response to the feeling tone itself, not to the object (sensory or cognitive) (Grabovac et al., 2011, p. 155). "With the awareness of any object, there is a concomitant feeling tone, which falls into one of three categories: pleasant, unpleasant, or neutral. [...] Due to the rapid and transient nature of these feelings, constantly arising and passing away, they often go unnoticed and can serve as the key trigger to a chain reaction of thought (including emotions) and actions that can lead to suffering." (p. 155)

15 Grabovac et al, drawing on Nyanaponika (2010), note that in Buddhist terminology these three insights are reffered to as (1) impermanence, (2) suffering, and (3) not-self.

ing reasons) that link core beliefs and specific thoughts. The concomitant feeling tone (pleasant, unpleasant, or neutral), in the sense proposed by Grabovac et al. (2011, p. 155), along with desires and volitions, further suggests that the thoughts in question reflect core beliefs. Acknowledging beliefs is the next step, which requires reflective analysis (through questioning), in order for beliefs to become conscious and explicit. This step goes beyond what is taught in mindfulness training.

Standpoints shaped by our acquired experiences tend to underlie claims charged with a strong emotional tone and tend to recur, continuously reinforcing the same evaluative stance. We can become aware of them through mindful meta-awareness, although our conversational partners often detect them even more easily. This is why it is sometimes harder to identify our own standpoint than to notice someone else's. For standpoints grounded in lived experience (e.g., the standpoint of an oppressed person or a migrant), their recognition largely depends on the mindful disposition — a long-term investment in understanding one's own experiences. Standpoints that stem from our social position — and are not accompanied by strong emotions — are often revealed through deliberate questioning, by considering factors such as our social status, profession, and other structural determinants of our position. Recognizing one's own standpoint is crucial because only then can we begin to understand its limitations. The next step is to examine one's own point of view in relation to alternative perspectives — a practice that Richard Paul (1981, pp. 3, 4) identifies as the essence of *strong sense* critical thinking. In contrast, merely evaluating arguments that stem from one's own perspective reflects *weak sense* critical thinking. We will return to the "ability to reason in more than one point of view", as Paul puts it (p. 4), through the topic of openness to others' points of view in subsection 4.4.

A meta-analysis of 21 morphometric neuroimaging studies published by 2014, which included approximately 300 participants in total, concluded that several brain regions show consistent structural differences[16] in individuals who meditate. Among them are regions that are key to meta-awareness and introspection, such as the rostral lateral prefrontal cortex (RLPFC/BA 10) (Fox et al., 2014, p. 64). However, as we presented, mindfulness meditation increases meta-awareness in terms of its quality. When we speak of meta-awareness in mindfulness practice, it is not expected to be active at all times, but rather to emerge at specific stages of practice. The mindfulness approach balances between non-reflective experiential consciousness and reflective awareness in the form of

16 They do not claim a causal relationship between meditation practice and structural brain changes, as they argue more research is needed to determine this.

mindful meta-awareness, with different degrees of engagement depending on the stage of practice. In the early stage, such as during focused attention or initial open monitoring, one is primarily sustaining attention and aware of a chosen object. Later, attention shifts to the flow of experience and includes awareness of all mental events as they arise and pass in the stream of consciousness (e.g., in open monitoring meditation; see Lutz et al, 2008). It is in this phase that insights associated with meta-awareness become more frequent.

4.3. Cognitive Flexibility Improved by Deautomatization

Cognitive flexibility refers to the ability to adapt one's cognitive processing strategies and behaviour to a given situation, and to cope with new and unexpected conditions within that situation. The data gathered from reviewing several empirical studies permits an examination of the effects of mindfulness training on cognitive flexibility through several theoretical constructs: deautomatizacion, thinking styles, meta-awareness, emotion regulation and tolerance of uncertainty. In this section, I will focus on the first two, since the remaining constructs are addressed in other chapters, and their relevance can be implicitly inferred.

Moore and Malinowski (2009) propose that the mechanism underlying the development of cognitive flexibility involves the intrinsic effect of the deautomatization of cognitive processes, which occurs through the reinvestment of attention. Moore and Malinowski build on classic cognitive psychology findings (Shiffrin & Schneider, 1977; Spelke et al., 1976) to theorize that cognitive processes can be categorized as controlled or automatic, with automatic processes operating independently of attention and without taxing short-term memory capacity. Some processes may become automatic through practice or repetition, after which they are difficult to interrupt or prevent voluntarily. (p. 177). The rationale is that if automated processes act independently of attention, deautomatization is achieved by reinvesting attention in actions and behaviours, which is precisely what happens in mindfulness training. The effect is stronger because the attention that is reinvested is non-reactive – for example, because it enables greater tolerance of negatively appraised aspects of experience, which might otherwise provoke superficial or distorted judgments of the situation. Findings from the Stroop test performance[17] in meditators suggest

17 The participant must inhibit the automatic tendency to read the word (e.g., 'blue') and instead name the color in which the word is printed (e.g., red). Demo of test: https://www.psytoolkit.org/experiment-library/experiment_stroop.html (accessed January **27, 2025).**

that cognitive processes which have become automatic can be brought back under cognitive control, and that responses which were previously automatic can be interrupted or inhibited (pp. 178, 182). Moore and Malinowski explain cognitive flexibility through both increased attentional capacity (as measured by the d2 test) and enhanced inhibition of automatic responses. Inhibiting automatic reactions creates the opportunity for a more objective perception of the situation and the emergence of new, more adaptive responses. Additionally, mindful meta-awareness aims to reach "reflexive" awareness of phenomenal field of each experience and to recognize the nature of one's emotional and cognitive patterns (Lutz et al., 2008, p. 162). While attentional control and response inhibition support cognitive flexibility by allowing for suppression and perceptual clarity, they do not encompass the generative dimension — the ability to create new ideas and strategies.

Ellen Langer uses a fitting phrase to describe how attention activated in a mindful state *counteracts, with a mindful outlook,* the processes that lead to mindless behaviour. A mindful outlook on the present moment generates new distinctions, new connections, new categories, inferences, and decisions. Eugene Gendlin sees thinking as a process that can either bring forth something new from experience or remain confined within already-established boundaries. "Logical thinking stays within whatever 'conceptual boxes' it starts with." (Gendlin, 1981, p. 57) Ellen Langer analyzes how automatic behaviour has much in common with categories and distinctions formed in the past – whether based on one's own experience, on socially accepted knowledge, or on the mere following of instructions and procedures (1989). Such categories, distinctions, and procedures may not make much sense in the present moment. Langer's experiments point to transformation from rigidity to flexibility in individuals who completed mindfulness training.

Divergent thinking refers to a type of cognitive process that generates multiple alternative solutions. Among the dispositions for critical thinking, Ennis lists *seeking alternatives* (to hypotheses, explanations, conclusions, plans, and sources) and emphasizes the importance of being *open to them* (Ennis, 1996, p. 9). According to Colzato et al. (2012), divergent thinking "allows many new ideas to be generated, in a context where more than one solution is correct" (p. 1). This type of thinking is often contrasted with convergent thinking, which is "considered as a process of generating one possible solution to a particular problem. It emphasizes speed and relies on high accuracy and logic" (p. 1). Researchers use the Alternative Uses Task (AUT) to assess divergent thinking, and their findings suggest that mindfulness enhances this capacity. As Colzato et al. write:

"Divergent thinking (as assessed by the AUT) would be likely to require or benefit from a control state that provides a minimum of top-down control and local competition, so that the individual can easily and quickly 'jump' from one thought to another in an only weakly guided fashion" (Hommel, 2012; Hommel et al., submitted; Colzato et al., 2012, p. 2).On the other hand, they note that convergent thinking (which they assessed using the Remote Associates Task – RAT) "would be likely to benefit from a strong top-down bias, which would heavily constrain and direct the search process, and from strong local competition (as only one solution can be correct)" (p. 2).

Fabio and Towey assessed cognitive flexibility through the framework of thinking styles, proposing that it involves whole-brain functioning rather than the dominance of a single hemisphere. They distinguish three thinking modes: "logical-analytical thinking, characterized by a preference for systematic approaches (originally left-brain dominance); an holistic manner of processing information in a synthesized and intuitive way (originally right-brained dominance); integrative mode of thinking that implies the use of a dynamic and interactive processing (originally whole-brained)" (2018, pp. 75) Using the *SOLAT test*, they found that meditators typically use an integrative style of thinking, meaning they flexibly combine both styles depending on different situation (p. 81). In contrast, the non-meditator group showed a left-dominant thinking profile, reflecting the typical distribution found in the general population — that is, a preference for structured tasks solved systematically. (p. 81) Among the brain regions that show consistent structural differences across various morphometric neuroimaging studies are regions key to intra– and interhemispheric communication — namely, the superior longitudinal fasciculus and the corpus callosum (Fox et al., 2014, p. 64).

4.4. Intellectual Courage and Open-Mindedness Improved by Emotional Regulation

Paul and Elder define intellectual courage as a metacognitive demand: "Having a consciousness of the need to face and fairly address ideas, beliefs, or viewpoints toward which we have strong negative emotions and to which we have not given a serious hearing. This courage is connected with the recognition that ideas we consider dangerous or absurd are sometimes rationally justified (in whole or in part) and that conclusions and beliefs inculcated in us are sometimes false or misleading." (Paul & Elder, 2020, p. 24)

This subsection explores the emergence of this "courage to explore complex or unpleasant truths" from the perspective of mindfulness. Rather than interpreting intellectual courage solely as a product of deliberate reasoning, it may also arise as an inherent outcome of the cultivation of a quality of mindful awareness that "views its object (sensory or cognitive) with neither attachment nor aversion", known as equanimity (Grabovac et al, 2011, p. 159). Grabovac and colleagues describe this quality of awareness as a "balanced state of mind" in which an equal interest is taken in all objects, regardless of whether they have a pleasant, unpleasant, or neutral concomitant feeling tone. Sensory or cognitive objects continue to be experienced as pleasant, unpleasant, or neutral (p. 157), and meta-awareness registers those concomitant feeling tones. However, one must release them rather than allow a habitual reaction (attachment or aversion) to unfold. These reactions typically manifest as subsequent mental events such as thoughts, memories, or emotions that arise quickly after a trigger (p. 155).

Although this is seldom discussed within epistemological frameworks, recognizing one's own emotional state in relation to a given topic (i.e., meta-emotional awareness) is crucial for understanding the thoughts that arise concerning the topic, identifying association triggers, acknowledging which beliefs enter the discourse, and noticing triggers that prompt maladaptive processing patterns. Our reasoning is influenced by emotions, needs, and desires. Equanimity contributes to improved emotional regulation, as non-reactive awareness reduces the need for cognitive effort in regulating emotions, creating a pause between experiencing and reacting. Heine and Dufner's (2025) study suggests that we are often unaware of our emotional state because there is a gap between spontaneous facial activity—often preceding conscious emotional recognition—and later reflective awareness. In other words, in some situations, meta-emotional awareness does not occur due to a lack of meta-bodily awareness.

Among the most frequently mentioned epistemic dispositions across virtue epistemologists and critical thinking theorists is open-mindedness – this involves seriously considering points of view other than one's own, reasoning from premises with which one disagrees without allowing the disagreement to interfere, and remaining open to changing one's views when necessary (i.e., openness to revision) (Ennis, 1987, p. 12; 1996, p. 372) The term openness highlights the quality of one's attitude toward new information and others' ideas. We often receive information without truly being open to it. For example, as parents or teachers, we may find ourselves trying to keep up with social media trends in the worlds of our children or students—but are we genuinely open to understanding them? The term *openness* may suggest acceptance, but it refers more precisely to the degree of one's involvement and willingness to engage with and understand new information. This also applies to the experience of one's own sense perceptions, emotions, and thoughts. The adoption of a positive or negative stance comes only after one has fully taken in the information.

Let us recall the findings from Sedlmeier and co-authors that the largest effect size is on interpersonal relationships. Mindfulness can positively impact the extent to which a person is able to listen and understand another person's perspective. Receptivity and responsiveness in interpersonal interactions foster high-quality intersubjective relationships and openness to others' views. An individual in a mindful state of awareness is non-reactive but receptive - not bound by habitual reactions — but responds with attention that follows whatever arises in the phenomenal field, thus displaying genuine responsiveness. This creates space for accurate, comprehensive, and subtle perception and understanding by introducing a pause between experiencing (perceiving) and reacting (e.g. thinking). Accurate perception of others – of their emotional states and needs – is foundational for compassionate responses (e.g. active listening). According to Varela, Thompson and Rosch, mindfulness helps cultivate such a relational stance toward others—one that encompasses not only empathy but also a sense of responsibility leading to concrete actions (such as considering others' perspectives). They note: "Another characteristic of the spontaneous compassion that does not arise out of the volitional action of habitual patterns is that it follows no rules. It is not derived from an axiomatic ethical system nor even from pragmatic moral injunctions. It is completely responsive to the needs of the particular situation." (Varela et al., 2016, p. 248) Paul and Elder find that intellectual empathy, as an intellectual virtue, is strongly correlated with the ability to accurately reconstruct others' viewpoints and reasoning (2020, p. 25). For them, narrow-mindedness is the opposite of intellectual empathy.

5. Limitations and Drawbacks of Mindfulness

In this paper, I explored the impact of mindfulness on information processing itself, covering both the advantages and disadvantages that a more mindful awareness introduces into cognitive dynamics. Mindfulness changes the dynamics of cognitive processes—sometimes by slowing them down, and other times by helping them become more direct and efficient. For example, in problem-solving situations a mindful person might refrain from immediately jumping to a conclusion, waiting to allow a fuller understanding of the problem. In conversation, instead of offering a quick reply, a mindful speaker might pause, sense their emotional state, and choose words more carefully. Mindfulness creates a delay between stimulus and response, interrupting impulsivity and allowing more deliberate reflection. In expertise (Langer's example), over-awareness of a familiar routine (like typing or playing an instrument) can interfere with fluid performance. It is not appropriate to enter a mindful state in every situation.

One reason is that becoming consciously aware or mindful can impair an expert's performance. If we know a task so well that we can perform it expertly, the steps may no longer be consciously accessible, and increased awareness might make us doubt our competence. When a frequently repeated task is slightly modified in an unusual way, a novice may perform better than an expert, as Ellen Langer (1989) notes.

A second reason is that mindfulness training cannot help a person think convergently; it is positively correlated with the development of divergent thinking, creating more hypotheses, and fostering creativity. According to the aforementioned study by Fabio and Towey, meditators are able to adopt a dynamic and integrative style of thinking. However, when it comes to using the logical-analytic mode of thinking, they will do so using skills not trained through the mindful state or whole-body information but rather through classical training in logic and scientific methodologies. Through mindfulness training, persons learn to be flexible, meta-aware, and to use their own emotions and thoughts that arise as additional information for questioning their own assumptions and point of view.

Third, staying in mindful state for too long can reduce productivity and efficiency. While mindfulness can improve focus and attention, staying in that state for too long can slow down decision-making and efficiency in tasks that require quick reactions. Ellen Langer claims that people who have undergone her mindfulness training are better at knowing what they want and making more intelligent choices, demonstrating a form of efficiency rooted in value-based and adaptive decision-making. However, excessive focus on 'here and now' can hinder long-term planning or the ability for strategic thinking.

In my opinion, the question is not whether it is worthwhile to cultivate a mindfulness disposition, but rather when it is appropriate to use and what it should be complemented with. Based on the discussion above, I emphasize the need to combine integrate mindfulness skills with analytical reasoning and proactive decision-making. My point is that, alongside the usual skills of scientists and philosophers—analytical and logical reasoning—the extent to which a person is capable of mindfulness can significantly influence comprehensiveness, accuracy, and subtlety of perception, cognitive flexibility, as well as openness to new hypotheses and alternative perspectives. Zagzebski emphasizes the importance of combining intellectual virtues, and of considering the relation between intellectual goods and other human goods. She cites Francis Bacon: "In fact, as is clear, the more active and faster a man is, the further astray he will go when he is running on the wrong road" *(Novum Organum,* trans, and ed. Urbach and Gibson, I.61), and offers the following putative case: A person who is intellectually courageous but lacks open-mindedness may be led further astray from the truth by his courage than he would have been without this virtue. [...] A person might be led into more false beliefs by the combination of her closed-mindedness and intellectual courage than she would have been if she had not had the courage." (Zagzebski, 1996, pp. 95–96)

Conclusion

The aim of this paper was to examine whether the training of non-reactive attention to the flow of present-moment experience significantly enhances epistemic dispositions—and in what aspects. I sought to clarify how perception and meta-awareness are altered, and how cognitive flexibility and emotional regulation develop through open-monitoring meditation. Using the language of epistemic dispositions, I aimed to explain why implicit beliefs and biases become more accessible, why the grasping of context and situation becomes broader and more accurate, why recognizing one's own standpoint becomes easier, why we become more open to others' perspectives, and why we become more willing to consider ideas that we dislike or find absurd. Several mechanisms contribute to these changes.

I concluded that non-reactive awareness, when trained, is highly beneficial because it promotes dispositions that are foundational and prerequisite for logical-analytical information processing. Especially helpful in this regard are the training of attentional functions (maintaining, allocating, and inhibiting), the non-identification with mental contents (thoughts, emotions, or impulses) in the flow of consciousness, and the non-reactivity to them, as this improves perception. Also, expanding reflective interest through meta-awareness in a mindful manner to all kinds of mental states and processes (cognitive, emotional, conative, physical, and behavioural), and tracking them equally through meta-awareness, proves essential for comprehensive perception.

Based on the presented findings, I argue that mindfulness practice contributes not only to social and psychological adaptability but also to more epistemically refined thinking— by increasing the capacity to perceive without bias, to metacognitively monitor one's reasoning processes, to prevent emotional reactivity from automatically influencing conclusions, and to reorganize one's interpretations of reality in light of new information. I suggest that enhancing reasoning capacity in this way is complementary to the traditional development of analytical capacities through the study of logic and methodology, and that together, they can make a scientist more epistemically competent.

These findings challenge our standard view of how epistemic competence is developed. It may seem unusual that a habit such as non-reactively attending to the unfolding stream of experience—or more simply, devoting greater attention to the present moment—can yield more successful epistemic reasoning. We typically imagine the scientist as predominantly cognitively active, whereas mindfulness training fosters a balance between non-reactive experiential consciousness and reflective awareness or deliberation. What becomes more important for scientists is the functioning of the whole person, through which they gain insight and self-regulatory access through meta-awareness.

A possible critique of this study is its use of exemplary presentation of studies[18] on which its conclusions regarding the influence of mindfulness training on epistemic dispositions are based. The presented body of research would benefit from supplementation with comparable studies that can further substantiate the claims. Moreover, future theoretical work could compare philosophical traditions that emphasize the role of present experience (e.g., phenomenology, hermeneutics, and pragmatism) with the mindfulness approach, in order to draw more sophisticated conclusions about the epistemological implications of cultivating an open attitude toward present-moment experiencing. I chose to focus on the mindfulness approach—although it is not traditionally philosophical—because its well-developed concepts, already explored in psychological literature, can serve as mediating tools in articulating its epistemological implications.

References

Baer et al. (2004). Assessment of mindfulness by self-report: The Kentucky inventory of mindfulness skills. Assessment, 11(3), 191–206.

Chin, J.M. & Schooler, J.M. (2009). Meta-Awareness. In W.P. Banks (ed.) *Encyclopedia of Consciousness, Vol 2* (pp. 33–41). Oxford: Elsevier.

Cohen, J. (1992). A power primer. *Psychological Bulletin* 112(1), 155–159.

18 I used findings from original psychological studies (Fabio & Towey, 2017; Zeidan et al., 2010; van Leeuwen et al., 2009; Moor and Malinowski, 2009; Colzato et al., 2012), psychological meta-analaysis (Sedlmaer et al, 2012), as well as neuroscientific meta-analysis of structural changes (Fox et al., 2014) about the effects of mindfulness training.

Colzato et al. (2012). Meditate to create: the impact of focused-attention and open-monitoring training on convergent and divergent thinking. *Frontiers in Psychology,* 3, 1–5.

Dunne, J.D., Thompson, E., & Schooler, J. (2019). Mindful meta-awareness: sustained and non-propositional. *Current Opinion in Psychology, 28,* 307–311.

Ennis, R. H. (1996). *Critical Thinking.* New Jersey: Prentice Hall.

Ennis, R. H. (1987). A Taxonomy of Critical Thinking Dispositions and Abilities. In J. B. Baron and R. J. Sternberg (eds.),*Teaching Thinking Skills: Theory and Practice,* W.H. Freeman and Company, 9–26.

Fabio, R.A., & Towey, G.E. (2018). Long-term meditation: the relationship between cognitive processes, thinking styles and mindfulness. *Cognitive Processing,* 19(1), 73–85.

Fox et al. (2014). Is meditation associated with altered brain structure? A systematic review and meta-analysis of morphometric neuroimaging in meditation practitioners. *Neuroscience and Biobehavioral Reviews, 43,* 48–73.

Gallagher, S. (2020). *Action and Interaction*. Oxford: Oxford University Press.

Gendlin, E. T. (1981). *Focusing.* New York: Bantam Books.

Grabovac et al. (2011). Mechanisms of Mindfulness: A Buddhist Psychological Model. *Mindfulness,* 2 (3), 154–166.

Heine, C., & Dufner, M. (2025). How aware are people of their current affect? A physiology-based investigation of affective awareness. *Emotion.*

Kabat-Zinn, J. (2003). Mindfulness-Based Interventions in Context: Past, Present, and Future. *Clinical Psychology: Science and Practice*, 10(2), 144–156.

Kabat-Zinn, J. (1994). *Whenever You Go, There You Are*. New York: Hyperion.

Kiken et al. (2015). From a state to a trait: Trajectories of state mindfulness in meditation during intervention predict changes in trait mindfulness. *Personality and Individual Differences, 81, 41–46.*

Kuburic, A. (2022). *Uloga iskustva i doživljaja u kritičkom mišljenju [The role of experience in critical thinking]* [Doctoral dissertation, University of Belgrade]. National Repository of Dissertations in Serbia. https://nardus.mpn.gov.rs/handle/123456789/22308

Langer, E. (1989). *Mindfulness.* Cambridge, MA: Perseus Books.

Lutz et al. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Science,* 12(4), 163–169.

Moore, A., and P. Malinowski. (2009) Meditation, mindfulness and cognitive flexibility, *Consciousness and Cognition,* 18: 176–186.

MacLean et al. (2010). Intensive Meditation Training Improves Perceptual Discrimination and Sustained Attention. *Psychological Science,* 21(6), 829–839.

Paul, R. W., and L. Elder (2020). *The Miniature Guide to Critical Thinking Concepts and Tools.* Lanham: Rowman & Littlefield.

Paul, R. W. (1981). Teaching Critical Thinking in the 'Strong' Sense: A Focus on Self-Deception, World Views, and a Dialectical Mode of Analysis. *Informal Logic,* 4(2), 2–7.

Sedlmeier et al. (2012). The psychological effects of meditation: a meta-analysis. *Psychological Bulletin,* 138 (6), 1139–1171.

Schooler, J. W., and I. B. Mauss (2010). To Be Happy and to Know It: The Experience and Meta-Awareness of Pleasure. In M. L. Kringeblach and K. C. Berridge (Eds.), *Pleasure of the Brain.* Oxford: Oxford University Press, 244–254.

Smallwood et al. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review*, 14(3), 527–533.

Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1.* New York, Oxford: Clarendon Press, Oxford University Press.

Sosa, E. (2017). *Epistemology.* Princeton: Princeton University Press

Sternberg, R. J. (2000). Images of Mindfulness. *Journal of Social Issues,* 56(1), 11–26.

Thompson, E. (2015). *Waking, dreaming, being.* New York: Columbia University Press.

Turri, J. (2015). Unreliable Knowledge. *Philosophy and Phenomenological Research,* 90(3), 529–545.

van Leeuwen et al. (2009). Age effects on attentional blink performance in meditation. *Consciousness and Cognition,* 18(3), 593–599.

Varela et al. (2016) *The Embodied Mind: Cognitive Science and Human Experience* (Rev. ed.). Cambridge, MA: MIT Press.

Zeidan et al. (2010). Mindfulness meditation improves cognition: Evidence of brief mental training. *Consciousness and Cognition,* 19(2), 597–605.

Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry Into the Nature of Virtue and the Ethical Foundations of Knowledge,* Cambridge: Cambridge University Press

https://doi.org/10.18485/baf.2025.2.ch7

Ognjen Milivojević

BACKGROUND AND ABILITIES AS DISPOSITIONS

Abstract: This paper examines the connection between John Searle's concept of 'Background,' also known as background powers, and the philosophical notion of dispositions. Searle defines the Background as a set of non-representational, pre-intentional capacities that enable intentional actions. Dispositions, on the other hand, refer to inherent properties that predispose entities to behave in specific ways under certain conditions. This study argues that Searle's background powers can be interpreted through a dispositional framework. My argument in this paper is that since background powers are (fundamental) abilities of the mind and abilities are dispositions, background powers are, therefore, dispositions. The paper builds on Vetter's critique of two dominant dispositional models of abilities and offers a revised understanding that aligns with the dispositional framework. This interpretation has significant implications for the philosophical understanding of abilities and Searle's concept of the Background.

Keywords: Background, dispositions, abilities, John Searle, Vetter, pre-intentional capacities, dispositional framework

Introduction

In this research paper, we will explore the connection between Searle's 'Background,' also known as background powers, and the philosophical concept of dispositions. John Searle introduces the 'Background' as a set of non-representational, pre-intentional capacities that underpin intentional actions. Dispositions refer to inherent qualities or properties that lead entities to behave in specific ways when certain conditions are met. The central question of this study is whether Searle's background powers can be understood through a dispositional framework. In other words, we ask whether background powers could be defined as dispositions.

My argument is as follows: since background powers are abilities, more precisely, fundamental abilities of the mind, and since abilities are dispositions, background powers are dispositions.

That background powers are abilities does not need to be specifically argued beyond analyzing the concept of the Background. The challenge lies in providing a dispositional analysis of ability. In this, I will build on Vetter's critique of two dominant dispositional models of ability. She provides strong counterexamples to these models and forces the dispositional analysis of abilities to either be abandoned or find a third solution while keeping the dispositional framework.

This study is divided into four parts. The first part presents and expands on Searle's conception of the Background. The second part contains a surface-level analysis of the notion of disposition. I will not delve deeply into dispositions but will adhere to a working definition around which there is a consensus among philosophers. The third part presents Vetter's criticism of two dominant dispositional conceptions of ability. Finally, in the fourth part, I offer my understanding of ability as a kind of disposition. Judging that the definition I provide is adequate, I conclude that background powers can be understood as dispositions since they are also abilities.

The significance of the considerations I make here is twofold. It informs the debate about the interpretation of abilities. As Vetter notes, the concept of ability is insufficiently researched, yet it is taken as a primitive concept in many philosophical fields. Defining ability may contribute to solving problems in the domains where this concept appears. In her words:

> "...abilities are invoked all over the place in philosophical theorizing: in accounts of concepts and of the mind itself (Millikan 2000; Kenny 2010), of conceivability (e.g., Yablo 1993), of agency (Mayr 2011) and activity (Groff 2013), of weakness of the will (Smith 2003), of omnipotence (Hoffman and Rosenkrantz 2012), of qualia (Lewis 1990), of empathy (Stueber 2006), of affordances (Scarantino 2003), and in non-virtue-theoretic accounts of knowledge (Hyman 1999, 2010) and its characteristic value (Carter et al. 2013), to name but a few. Most, though not all, of these use abilities as unexplained explainers." (Vetter, 2003, p. 216)

Second, Searle's notion of the Background is philosophically interesting in itself. Although it is well covered by Searle, he doesn't provide a dispositional reading of the Background. Here, I will attempt to do that. If successful, it will add to our understanding of this important notion.

1. Background Powers: Definition and Types

Searle's concept of background powers is fundamental to understanding how humans engage with the world automatically and seamlessly. These background powers refer to the set of non-representational, non-conscious capacities that underpin our intentional actions and thoughts. Unlike intentional states, which are about something specific, background powers operate behind the scenes. They include skills, habits, and assumptions we rely on without thinking. For example, when we speak a language fluently, we don't actively reflect on grammar or syntax—this ability is part of our Background.

Searle divides these background powers into two categories: the deep background and the local background.[1]

The deep background refers to universal, biologically ingrained capacities shared by all humans. These include basic abilities like recognizing objects, walking, or interpreting facial expressions. These capacities are essential for survival and interaction, operating in a pre-reflective way that we never consciously consider. No matter where you come from or what your cultural background is, these deep background powers are present and active.

The local background, by contrast, consists of culturally specific practices and skills shaped by one's social environment. These include the ability to navigate social norms, understand cultural symbols, or practice local customs. For example, using chopsticks versus a fork, or interpreting cultural gestures like a handshake or a bow, are part of the local background. These skills are still automatic and non-conscious in their application but are learned through social experience and shaped by cultural context. The local background, therefore, differs across societies and is crucial for navigating particular cultural landscapes.

While background powers operate automatically, they provide the necessary foundation for intentional states. Intentional states, such as beliefs, desires, and intentions, are representational—they are about something specific and are often consciously directed toward a goal or object. For instance, when you decide to greet someone, your intentional state involves the belief that greeting is appropriate in this context and the intention to perform the greeting. However, the actual mechanics of greeting—knowing how to smile, shake hands, or bow—are governed by your background powers, particularly your local background.

Intentional states rely on the Background to function smoothly. Without the Background, intentional states would be ineffective or impossible to carry out. For example, forming the intention to have a conversation presupposes that you have the background ability to understand and produce speech. Moreover, while background powers enable intentional

1 Searle introduces this distinction in his book *Intentionality: An Essay in the Philosophy of Mind*. In his numerous subsequent works where he uses the term Background, the distinction is not made explicit, although it is implicitly acknowledged.

actions, they are non-representational—they do not involve conscious reflection or thought about how these actions are performed. Instead, they work in the background, as their name suggests, providing the necessary skills and know-how for intentional states to be enacted.

2. Basic Analysis of Dispositions

Now we turn to dispositions. Dispositions refer to qualities or properties that predispose an entity to behave in certain ways under specific conditions. For instance, fragility is a classic example of a disposition—an object is said to be fragile if it tends to break when struck.[2]

Dispositions could be contrasted with occurrent (or categorical) properties. Occurrent properties would be properties that manifest at a given moment. They represent the current state of an object or system and are directly observable. For instance, the temperature of a cup of tea, the redness of a rose, or the velocity of a moving car are all occurrent properties. These properties are active and present at the moment of observation and do not depend on external conditions for their manifestation. Unlike dispositions, which may never be actualized unless the right conditions arise, occurrent properties are present and observable right now. In other words, dispositional properties (like fragility) refer to potential behaviors or outcomes that an entity may exhibit under the right circumstances, while occurrent properties are about what is actively happening in the current moment. An object can be fragile (a disposition) without actually breaking, but it can only be at a specific temperature (an occurrent property) if that temperature is observable now.

Dispositions are typically described with two components: the stimulus condition (in our example, the striking of the object) and the manifestation (in our example, the breaking of the object). The stimulus condition is what triggers the manifestation to occur.

Specifying a particular disposition, like fragility, involves specifying its stimulus condition(s) and its manifestation. Similarly, to determine a kind of disposition—such as determining what kinds of dispositions are abilities—the kind of stimulus condition and the kind of manifestation must be provided.

---

2 Although definitions of dispositions remain philosophically contested, the one adopted here reflects a broadly accepted starting point. As Crane notes, even a preliminary account can serve specific purposes without settling all theoretical questions (Crane, 1996, p. 1). The aim of this paper is not to provide a comprehensive theory of dispositions, but rather to offer a definition of abilities understood dispositionally.

The contrast drawn here between dispositions and occurrent properties follows a widely accepted distinction in the metaphysics of dispositions (Mumford, 1998; Molnar, 2003). Occurrent properties are those instantiated and observable at a specific time (e.g., velocity, temperature), while dispositions are modal properties that may never manifest unless suitable conditions arise (e.g., fragility, solubility). While it is possible to analyze dispositions in terms of the distinction between possession and manifestation, the occurrent/dispositional contrast provides a clearer ontological grounding for analyzing abilities as capacities that may or may not be exercised, depending on circumstances. This treatment is consistent with the purpose of the paper, which is to argue that abilities—and thus background powers—are a species of disposition.

3. Vetter's Criticism of Major Dispositional Readings of Abilities

Having gained basic insight into the notion of dispositions, we now move to Vetter's criticism. Vetter critiques two major approaches to defining abilities through dispositional analysis: (1) abilities as dispositions to do what one intends to do, and (2) abilities as dispositions to successfully do what one already does. Articulated through the components described above, according to the first understanding, the ability to A (where A is a simple or complex action) is a disposition whose stimulus condition is the intention or conscious attempt to do A, while its manifestation is (the execution of) A. According to the second understanding, the ability to A is a disposition whose stimulus condition is (the execution of) A and the manifestation is the successful, i.e., good, performance of A.

Both of these approaches, she argues, fail to account for the complexity of abilities. The first view, which defines abilities as dispositions to act according to intention, oversimplifies the way in which abilities function. It ignores abilities that may not be linked to explicit intentions, such as innate or automatic skills like motor activities or creative processes, which can occur without the agent forming a specific intention to act. It is well established that these types of abilities cannot be performed when we (consciously) try to perform them. As Vetter notes:

> "On the first account, we are to include abilities for complex actions but not those for their components; we are to include my ability to write this paper, but not my romantic poet's ability to write a beautiful poem; and we are to include the ability to dance, but not the ability to dance absent-mindedly. It is hard to argue about intuitions of gerrymanderedness, but I find this result unacceptable." (Vetter, 2019, p. 217)

The second view, which ties abilities to the disposition to succeed, also faces a refutation as a universal account of abilities. It does identify abilities where the success conditions are sufficiently distinct from the basic exercise of the abilities, but it fails to include examples where the exercise of ability coincides with its success criteria. These include simple motor abilities at the least. For example, in the case of the ability to bend one's leg, the success condition, which is the manifestation component of the disposition, is identical or almost the same as its stimulus condition:

> "The same kind of consideration would seem to apply quite generally to simple motor abilities: the ability to move my eyes, bend my leg, wiggle my foot, and so on. All of these performances are such that to perform them at all is already to perform them successfully. I simply have not bent my leg if my leg is not (at least a little) bent. Thus the ability to bend my leg, by (D2), should be the disposition to bend my leg well enough if I bend it at all, which is nothing but the disposition to bend my leg if I bend my leg; and so forth." (Vetter, 2019, p. 216)

Put simply, the 'success' account presupposes the distinction between a basic exercise of ability (stimulus condition) and doing it well (manifestation), yet this distinction is canceled when performing the simplest abilities, making it inapplicable to such cases. A universal account of abilities should be able to explain why we also call those cases abilities.

In summary, Vetter's critique shows us that dominant conceptions of the dispositional understanding of abilities are restrictive, each in its own way. They fail to include in the category of ability cases that are obviously exercises of abilities.

Vetter's critique emphasizes the need for a more nuanced understanding of abilities—one that does not reduce them merely to intentions or success but allows for dormant capacities that may not be explicitly linked to either. This is particularly important when analyzing Searle's Background, which consists of foundational abilities and capacities that often operate without conscious intent or awareness, and whose success criteria and exercise seem indistinguishable.

4. Abilities as Dispositions

As mentioned earlier, to determine what kinds of dispositions are abilities, we must determine the nature of their stimulus condition as well as their manifestation.

Acknowledging and building on Vetter's criticism, I propose that abilities, and by extension the Background, can be understood as dispositions whose manifestation is positively adaptive, extraphysiological behavior, and where the stimulus condition is an object, event, or environment that prompts the individual to exhibit the said behavior, or more precisely, an opportunity.

This definition aligns with two central features of abilities: (1) that abilities are functional capacities that, when exercised under suitable conditions, typically support the agent's engagement with the environment, and (2) that abilities are not simply physiological or physical processes, but involve extraphysiological behavioral responsiveness shaped by context. [3]Although a physiological process accompanies and underpins an exercise of an ability, an ability involves the mental engagement of the actor. As the analysis of the Background suggests, as well as Vetter's considerations, the psychic aspect of an ability doesn't need to be directly observable by the actor; rather, it can operate below the threshold of consciousness.

In addition to being consistent with these intuitions, my dispositional account of ability is immune to the aforementioned counterexamples that Vetter has raised to the dominant dispositional models of ability described. Namely, I propose that for something to be an ability, it is necessary and sufficient to establish that a) it generally contributes to one of these three tasks: the survival and reproduction of the organism, and its harmonious integration with the environment, and b) that it is not a purely physiological, i.e., physical process. Aimless walking, creative artistic processes, simple unconscious motor actions, etc., satisfy these requirements to a greater or lesser extent.

The definition that I propose excludes, I argue, behavior with typically negative consequences as an exercise of an ability, as well as purely physiological or physical processes. We are reluctant to say that atoms have an ability to bond, that bacteria have an ability to metabolize glucose, that plants have an ability to perform photosynthesis, or that an animal has an ability to grow hair, regardless of the fact that these processes contribute to their survival, reproduction, or integration with their environment. We would use some other term to describe such properties where a "positive adaptation" is a candidate.

3 While the definition emphasizes the functional nature of abilities, it does not claim that abilities are inherently valuable or always lead to successful outcomes. This is not a normative claim but a structural one: abilities are capacities that tend to lead to functional engagement when environmental conditions permit. In this sense, an ability can exist even if it fails to manifest, due to constraints in the environment. Thus, the dispositional model adopted here does not exclude or contradict the view that ability is co-defined by agent and environment—it simply analyzes the structural form of those capacities.

Additionally, maladaptive behaviors, like hazardous gambling[4], are often called disabilities rather than abilities. For the most part, we don't consider them abilities precisely because they are not beneficial for the one who does them.

Temptations to consider such actions as abilities stem from the fact that many, perhaps all, maladaptive behaviors include sub-actions that are, in and of themselves, positive adaptations or lead to beneficial outcomes for the agent or organism. For instance, gambling when the odds are against the player consists of subtasks like walking, being still for a while, recognizing shapes and colors, speaking a language, and so on. These simple and more complex actions are, considered individually, abilities. However, the complex action of high-risk gambling, and similar activities, is not an ability due to no net positive adaptive effect.

We may say that such inferences, like inferring that high-risk gambling is an ability, represent a fallacy of composition. In this way, based on the positively adaptive value of the constitutive sub-abilities, it is concluded that the encompassing action is an ability. But, we have no right to conclude based on the fact that some parts or steps of a complex action are abilities that the whole series of actions as a collective can be considered an ability.

Conclusion

The goal of this research paper was to determine whether a dispositional reading of the Background can be obtained. I found this to be possible since background powers are abilities and since abilities can be interpreted as dispositions. The main takeaway from this paper is that background powers or, more broadly, abilities are dispositions to behave in a positively adaptive, not purely physiological or physical mechanical manner when opportunities arise.

References

Searle, John R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

4 By this i mean risk-taking activity whose expected utility is lower than abstaining from it. The image of a broke person playing roulette should come to mind. Therefore, not all risk-taking is considered a disability—only those cases in which the expected utility is lower than that of abstaining from it. For example, quitting one's job in the hope of finding a new, better one could be understood as an ability in certain contexts.

Searle, John R. (1995). *Construction of Social Reality*. New York: The Free Press.

Searle, John R. (2010). *Making the Social World: The Structure of Human Civilization*. New York: Oxford University Press.

Vetter, Barbara (2019). Are abilities dispositions? *Synthese* 196 (196):201–220.

Crane, Tim (1996). Introduction. In *Dispositions: A Debate*, by D.M. Armstrong, C.B. Martin, and U.T. Place, edited by Tim Crane, London: Routledge.

Mumford, Stephen. (1998). *Dispositions*. Oxford: Oxford University Press.

Molnar, George. (2003). *Powers: A Study in Metaphysics*. Oxford: Oxford University Press.

4. DISPOSITIONS AND ONTOLOGY

https://doi.org/10.18485/baf.2025.2.ch8

Una Popović and Srđan Šarović

DONKEY'S DILEMMA: VALUES OR VALOR?

Summary: This paper explores the process of creating artwork, with a particular emphasis on the artist's role in that process. The concept of the 'donkey's dilemma' is used to illustrate the artist's position when faced with the act of creation. Specifically, the artist must decide whether to adhere to the inherent principle of the artwork, ensuring that all stages of its creation are aligned with this internal logic, or to conform to an external principle, thereby adapting the artwork to fit preexisting frameworks. The inherent, inner principle guarantees a coherent relationship between the metaphysical essence of the artwork and its physical form, ensuring that this principle is clearly expressed in the final work. In contrast, the external principle represents external ideas, ideologies, or cultural contexts that assign value and meaning to objects considered art within a given context. The artist, like the donkey, must bear the burden of his choice, and the dilemma lies in deciding which burden to accept – a dilemma we argue is ultimately false. In either case, the artist remains inextricably bound to the moral and ethical consequences of his choice regarding both the artwork and the principle guiding its creation. Art created in accordance with an inner principle possesses intrinsic value – a valor derived from the work itself, grounded in its own internal logic. By contrast, art tied to an external principle lacks inherent value, as its worth depends on the value system the artist adopts. This paper further argues that such external value systems, including widely accepted aesthetic standards, are context-dependent and therefore relativistic. By contrast, the genuine disposition for an artwork to possess value resides in its ontology – that is, in its inner principle. Accordingly, the decision between an inner or external principle in the creative process is, fundamentally, a choice between authentic value – that is, the valor, and imposed perceptions of value shaped by societal constructs.

Keywords: artwork, creation, inner principle, external principle, valor, values.

In this paper, we will address the issue of creating artwork, with a particular emphasis on the artist's role within that process. Our position is grounded in artwork production; this paper results from thinking about art from within the experience of creating works of art.

The concept of the 'donkey's dilemma' in the paper's title is designed to illustrate the artist's position regarding the artwork he aspires to create. It draws a parallel between the artist's decisions and the symbolic figure of a donkey confronted with an impossible choice. In this case, the donkey serves as a metaphor for the artist; such an artist-donkey is facing a particular dilemma within the process of creating a work of art.

The dilemma exemplifies the dual options the artist is presented with while deciding the way and manner in which he will create the artwork. The options in question have to do with the principle according to which the artwork is to be created: the fundamental general idea defining the artwork's meaning, structure, and physical appearance.

The principle of artwork's creation can be either the inherent or the external one. The inherent principle of the artwork (as such) is devised in the artist's mind specifically for creating the artwork in question; the establishment of the principle is the response to the ineffable *poiesis*, which incites the artwork's creation. On the other hand, the external principle inhabits the realm of culture, usually in the form of some ideology or ideology-related ideas. Therefore, when creating according to the external principle, the artist is not responding to the artwork's own being, nor does he define the principle. In this case, the artist conforms to some existing idea, consequently adjusting his creation to it (both as a critic or as an advocate of appropriated external principle, i.e. discourse).

So, the dilemma of the artist is which kind of principle to choose – the inner and inherent principle of the artwork (as such) or the external cultural principle? In other words, the dilemma is either to rigorously follow what is given through *poiesis*, disregarding any dialogue with existing cultural frameworks, or to engage with such cultural frameworks, whether the artist intends to confirm or to subvert them.[1] In either case, the chosen principle will determine the entire process of artwork creation; including all the steps in the mind of the artist and all the steps of physical making the artwork.

In our opinion, such a dilemma should be presented alongside another problem: the choice between valor and values. The two dilemmas are interconnected and overlap in the sense that choosing valor involves

1 Although the division between inner and external principles might seem too strict, we argue this is not the case. The division underlines the possible ways for an artist to understand and conduct his work. Although many artists today choose to interact with cultural traditions or political ideas, we believe they are not restricted to such behavior. Finally, the idea of artists being a priori confined to existing (or traditional) cultural frameworks could also be considered a consequence of the postmodern ideology.

the selection of the inner artwork's principle, whereas choosing values is equivalent to the choice of an external principle. As we will argue, the donkey's dilemma is, in fact, a pseudo-dilemma.

In order to introduce, present, explain and resolve the donkey's dilemma, we will first discuss the concept of artwork's principle and its significance in the process of creating the artwork. Following that, we will address the two main types of artwork principles: the inner (creative) principle and the external (cultural) principle. As a result, the donkey's dilemma and the artist's position in creating the artwork will become evident. Such formal analysis will be further enhanced through questioning the ethical aspects of creating the artwork, i.e. by comparing the artist's decision between two principles to the choice between valor and values.

The Artwork's Principle

The donkey's dilemma is a metaphor for the artist's position in the process of creating a work of art. The artist's position is constituted by his relation to the emerging artwork: the artist has to make a decision about what kind of work he will create and in which way he will act to accomplish this. In other words, the artist decides upon a principle of the artwork – a principle which defines the artwork's meaning and being, while at the same time serves as a guideline for the process of artwork's creation. The artwork's principle, as said before, can be either the inner or the external principle.

To put it clearly, the concept of artwork's principle primarily refers to the artwork's being and ontological constitution. The principle determines all essential features of the artwork and provides its meaningful inner structure. In other words, the principle is the foundation of the artwork's constitution, defining its elements and composition in a formal and general way. Such formally defined meaningful composition is further embodied into a material object through the process of making the artwork; an artwork's physical appearance is merely a final manifestation of its principle. Therefore, artwork's ontology and being are defined metaphysically – before the artwork is realized as a physical and material entity (and before its perception by the observer).

To create the artwork, the artist acts according to some particular (inner or external) principle which provides the ontological structure of the work in becoming. The artist creates the artwork according to a particular principle whether or not he is aware of the principle he is following. In other words, the artwork's principle can be explicitly formulated before the making of the artwork, or it can implicitly guide the artwork's coming into being. In either of these two cases, the principle determines the outcome of acting according to it. Whether the artist is aware of the principle or not, he organizes the entire process of creation in alignment with (or under) that principle, so that the principle determines and regulates the artwork's production in all its constitutive steps.

In other words, for the artist, the principle is equal to *the artistic method* directing the entire way of his acting while creating the artwork (Šarović, 2021, p. 163). Therefore, the principle has two functions: first, it is the ontological foundation of the artwork, and second, in the very process of creation, the principle serves as an artistic method (guiding the process).

The artistic method should not be confused with the (technical) artistic procedures employed in the very act of making the work of art or with the use of diverse techniques in working with materials and media (i.e. with the methodology of technology used and the principles of media and materials). The artistic method (i.e. the principle) is defined before the formation of the artwork's material form. The artistic method directs the artwork's creation completely, including the making of the artwork as a physical being which is simply the concluding phase of the creation process.

An artwork can find its living space in any medium or it can be created using any technology whatsoever; it can present itself as a poem or as a painting, symphony, or an NFT (non-fungible token / crypto art), or be classified as a work of popular, applied, or high art (Grugan, 2009, p. 6). What makes it the work of art is none of the enumerated categories; it is the principle on which the artwork is based and which the artwork manifests. The principle as such is independent of artistic media, technologies, and procedures – so much so, that the creation according to a particular principle can be realized in any medium and through employing any technology and procedure. Moreover, the creation can be realized in multiple media (McIver Lopes, 2014, p. 195–196); an artwork's principle can equally manifest itself in several media, through the employment of various technologies and procedures.

The principle directs the process of creation by providing an order of formal steps to be followed. Firstly, the principle (i.e. the artistic method) determines all the steps in the creation of the artwork prior to its physical making – specifically, the steps in creating the metaphysical artwork and its (metaphysical) meaningful composition. Secondly, the principle (i.e. the artistic method) determines the choice of medium (or several media), the exact procedures and the way of usage of artistic skills within that medium, as well as all particular steps in making the artwork as a physical being (Šarović, 2021, p. 164–165). The process of physical making of the artwork is organized according to the principle so that the physical composition and appearance of an artwork are the embodiment of its metaphysical being and composition. Consequently, the artwork's principle is manifested in the created artwork's appearance (perceivable by the senses).

Therefore, the artwork's perceivable appearance is not all there is to be seen (Dillenberger, 2004, p. 220; Henry, 1973, p. 438). There is also the essential, invisible side of the artwork – its meaning, shining through its perceivable and physical side. The invisible meaning of the artwork is equal to the metaphysical (invisible) world constituted according to the artwork's principle and before the artwork's physical realisation. The visible (and perceivable) appearance corresponds to the invisible meaning of the artwork because both are created according to (and related through) the same principle.

To put it differently: the artwork presents and communicates its state of being through its composition and appearance. This means that the overall reception of the artwork – grasping both its physical appearance and its metaphysical being – depends on and is essentially regulated by the artwork's principle. Since the artwork's principle is present and manifested in all aspects of the artwork, including the ones that are perceptible by the senses, the audience is also presented with the principle regardless of whether or not the principle was explicitly communicated by the artist. If the principle (i.e. artistic method) is adequately articulated and communicated, this will prevent misunderstandings regarding possible interpretations of the artwork, clearly showing which interpretations acknowledge the artwork's being (and meaning) and which ignore it. If the principle is not explicitly formulated and communicated, the recipient has to infer the principle, given the artwork's physical being and appearance.

Finally, the artist can attempt to produce an artwork without any principle. This occurs when the artist is focused solely on the physical making and appearance of the artwork, thereby emphasizing artistic skills and technological proficiency. In this situation, the artist attempts to combine the artwork's elements into a coherent and harmonious (physical) composition without any specific meaning guiding his work. However, since there is no meaning to his work, the outcome manifests and communicates (literally) nothing meaningful (Beardslay, 2019, p. 24).[2] Without the principle involved, such an attempt will fail as artwork, although the skill

2 If the artist deliberately chooses to create a meaningless artefact, such a choice would represent a specific principle of creation. However, such a principle could only be an external principle.

in using materials, presented with the created product, can be highly perfected.

This proves that artistic skill and technology alone are neither equivalent to, nor sufficient for the creation of art, and explains the observer's confusion when presented with such a product devoid of any principle. The observer can resolve his confusion by imposing some external meaning on the artefact by inventing a (new) principle as a framework for its interpretation. In this case, the observer is taking the position of a conceptual artist: he converts a meaningless work into a seemingly meaningful one by utilizing such a product (Osborne, 2000, p. 88–89).

The Inner Principle and the External Principle

When the artist creates an artwork according to some principle, the principle can be either the inner creative principle (of the artwork as such) or the external principle. Artwork with an inner principle is autonomous; it has its own essential constitution and meaning, independently of any further interpretation and explanation (Lamarque, 2013, p. 61). In contrast, artwork with an external principle is not autonomous (Roberts, 2023, p. 71–73). Therefore, such artwork is dependent on subsequent interpretations, which provide the meaning of the artwork by construing it from a particular viewpoint – by clarifying the relations between the created work, its (external) principle, and the culture into which the work is introduced (Danto, 1973, p. 15; Young, 2001, p. 29).

The inner principle. The creation and the meaning of artwork with an inner principle are inherent to that artwork. The artwork's creation begins with a certain something that wills itself to be, which comes upon the artist and cannot be ignored; this part of the creation process is traditionally known as *poiesis* (or inspiration, etc.) (Gaut/Livingston, 2003, p. 13–14). The inner principle is formed in the mind and soul/spirit of the artist as an answer to the question of how to adequately respond to whatever came through *poiesis* and to bring it forth into physical existence as a created work of art.

Therefore, the artist participates in the artwork's autonomous coming into being: the formulation of the artwork's principle is the first step of such participation, but the conditions of possibility for the inner principle to be formulated come through poiesis. The inner principle vouches for the autonomous (self)organisation of the artwork, and enables the artist to remove himself from the artwork's ontology; otherwise, the artwork has no being without the artist. Finally, if properly followed, the inner principle guarantees a consistent relationship between the artwork's metaphysical being and its physical appearance, ensuring that the principle is clearly manifested in the completed artwork.

When an artwork is created according to its inner principle, the artwork directs its own creation from within and autonomously organizes itself into a specific visibility and appearance. However, the artwork's elements and composition are not arbitrarily chosen. On the contrary, the artwork's inner principle directs the creation process towards particular elements in a particular order, so that there is an inner necessity manifested in the completed composition. In other words, artwork's inherent meaning is consistently conducted from an abstract (metaphysical and conceptual) composition to some particular physical composition and appearance. In such a process, the role of the artist is to remove anything that would prevent or constrain the artwork's autonomous self-organization. To do so, the artist has to recognize, comprehend, fashion and rigorously follow the principle (i.e. the method) up to the point the artwork has reached its completion.

Such an autonomous artwork is the embodiment of its inner principle, and so the principle is manifested through the artwork's perceivable appearance. Therefore, the artwork with an inner principle is the only point of reference to which we should turn to understand what the artwork is, how it is constituted, and what it means – to comprehend it ontologically and aesthetically. The autonomous artwork presents itself in a manner that escapes full rational explanation but enables complete comprehension of what it manifests and communicates.

The external principle. In contrast, the external principle is not derived from the artwork's own being (Thompson, 1999, p. 260). The external principle is constituted independently of the artwork and only then applied as a prescribed canon of its production. Once such an artwork is created, the external principle also functions as an established criterion of the artwork's (aesthetic) evaluation within the culture.

The external principle refers to any preexisting idea, ideology, or cultural standpoint that establishes the value framework against which some meaning and role are being ascribed to objects regarded and acknowledged as art within that context. In other words, in every culture, there are some ideas and values which determine the criteria for what is to be acknowledged as an artwork in that culture (Lamarque, 2010, p. 172–174; Carroll, 2012, p. 123–124). Being acquainted with such cultural norms, the artist can choose to accept them and turn them into a principle according to which he will act while creating the artwork. Such an external principle will then determine the steps in the process of the artwork's creation and also the entire constitution of the final product (Hemingway, 2017, p. 35). Therefore, formally speaking, the external principle becomes a sort of pseudo-inner principle of such work; it is an internalized external principle – the external-inner principle.

The external principle is not inherent to the artwork, but rather borrowed from some cultural ideology or paradigm; as a result, the artwork is an instantiation and embodiment of the paradigm from which its (external) principle originates.[3] In creating an artwork, the artist directs the product along cultural (or sub-cultural, or contra-cultural) norms and standards, and so the governing cultural norms operate as a regulative framework for the production (and reception) of his work. In this situation, the artwork is projected according to some prescribed purpose and imposed meaning; thus, the resulting outcome is more appropriately referred to as an artefact rather than an artwork (Beardslay, 2019, p. 28).

The creation of such a work of art is adjusted to external criteria and purposes, and so the artefact's meaning is also determined externally. Such an artefact is not autonomous; its meaning and (aesthetic) value are transferred from the paradigm of its external principle (Groys, 2013, p. 13). In that case, the artwork/artefact will be ascribed various interpretations derived from the external principle of its making, and such interpretations will pose as its pseudo-meanings (Carroll, 1994, p. 11–13; Crowther, 2013, p. 14).

The Dilemma?

Hence, the dilemma: while creating the artwork, the artist must choose between the inner creative principle and the external cultural principle. In other words, the artist-donkey must decide which burden to carry. Opting for the external principle, the artist decides to play along with the rules of an already established system, hoping for the benefits of being acknowledged and verified by the cultural institutions. The artist who adheres to such norms may hope to enjoy visibility, validation, and success within the existing cultural framework. Opting for the inner creative principle, the artist risks that his work will not be recognized or properly evaluated within society. However, in that case, the artist chooses the artwork as such and preserves its integrity. It seems that both options are equally reasonable and equally valid; and so, the donkey (artist) is stuck.

However, the donkey's dilemma is a false one. The choice between the inner and the external principle is not a choice between two equal options.

3 A clear example of this would be any politically engaged art: Social realism, Feminist art, etc.

The 'donkey's dilemma' is more than a mere choice between two burdens; the choice comes with profound consequences regarding the artist's and artwork's role in society. This ethical dilemma shows that artwork creation is, undoubtedly, an act with significant moral consequences regardless of context, historical circumstances, and prevailing ideologies in the culture.

The donkey's dilemma – whether to choose the inner or the external principle of creation – is a matter of evaluation of principles. Principles can be evaluated according to the worlds they constitute and propose, and which are manifested through the works created according to those principles. The world that can be seen through the artwork clearly shows what kind of principle it is based on – namely, whether the order established under the principle is meaningful and harmonious or not, and which values such an order proposes. The same holds true for the artefact, as it also demonstrates its principle and the underlying value system (i.e. the ideology of its external principle). Artwork with an inner creative principle brings forth the (artwork's) world constituted according to that principle, whereas artefacts made according to an external principle produce nothing of importance. It simply supports and manifests some ideology.

Consequently, the artefact lacks the valor, i.e. the autonomous value. The value of an artefact is solely determined by the system/paradigm to which the artist subscribes and from which its principle originates (Carroll, 2021, p. 64). The degree to which a work successfully supports its paradigm is the criterion of evaluation; hence, artefacts may have less or more value relative to one another depending on how effectively they represent and reinforce their associated paradigm/system. In contrast, creating according to an inner principle results in an artwork with intrinsic worth (valor) – a 'value' that is immutable and inherent to the artwork. Such artwork is autonomous regarding cultural (or sub/contra cultural) paradigm: autonomous artwork can be a critique of the cultural paradigm not as a mirror of society but as a normative or corrective proposal to society.

Works created according to an inner principle cannot be classified as more or less valuable. Rather, such works transcend any form of evaluation based on external value systems; they are subjects for evaluation solely if the evaluation is based upon the soundness of the world proposed through the artwork itself. The world of the artwork constituted through its inner principle, is either possible, necessary, coherent and consistent, or impossible, unnecessary, incoherent and inconsistent (and therefore of no significance). There is no external value system that applies to the sound world of the artwork; an external system is just that – something not intrinsic to the artwork. In other words, genuine works of art are invaluable and of inestimable worth. External values, including observers' accepted aesthetic values, are paradigm-dependent and hence relativistic.

To put it in another way, the artist creating the artwork must make a conscious and mindful choice, because his choice between the principles (whichever he chooses) reflects his own moral autonomy or lack thereof. Even if the artist is truly supporting some ideology or trying to subvert it, choosing the external principle is equal to the relativistic selection of the values for one's own benefit; including the mercantilist one, which is, in one way or the other, the basis for all other external values (Groys, 2013, p. 18–19). In this, the artist is free to act according to his own interests relative to the context; he can adhere to one ideology at a time, and then switch to the other with no particular repercussions (Gover, 2018, p. 13–14). However, such behavior is anything but moral. Even if this is a deliberate and conscious choice, it is a Sophistic response to a dilemma, presenting a wrong choice as a legitimate one.

On the other hand, if an artist deliberates on an inner principle of artwork's creation, he has chosen valor over (all external) values. The valor belongs to the world proposed with the created artwork as a sound, coherent and harmonious whole. It implies an autonomous choice, regardless of all context-related value systems, relativistic interests, and cultural norms. It also demands an *arete* to acknowledge and confirm the function of the artwork within the society. To accomplish this, the artist has to be fully aware of the (metaphysical) world that his work will bring into being. The artist must consider his position, his approach to creation and the principle he adheres to. In other words, the artist, as the accountable one, has to be a philosopher – constituting and evaluating an ontological model.

Concluding remarks

The donkey's dilemma results from an imbalanced relationship between the artwork (and the artwork's creation) and the society to which the artwork is introduced. At its core, the artist's dilemma is about the moral and ethical responsibility regarding the artwork he creates. The true question here is whether the culture should define and determine the boundaries of the artwork's being, meaning and role in the society. Or, perhaps, should the artwork provide a model for the organization of society and culture? In creating the artwork, the artist is faced with a choice between two possible worlds: the one in which the meaning of the artwork is marginalized and essentially (constitutively) insignificant, being in the function of an already existing societal order, and the other in which the artwork has a crucial role in the way the society organizes and understands itself, providing the model for a different (and better) society.

In other words, if there is such a donkey's dilemma, that means that the artist attempts to absolve himself of moral responsibility for the artwork he creates and for the way his work will impact society. The artist-donkey hopes to lighten the load, and therefore he is tempted to conform to the preexisting cultural and societal order. The burden of responsibility is then transferred to the ideology or the paradigm to which the artist has subscribed and from which he has borrowed the (external) principle of his artefact. However, by choosing to create the artwork according to an external principle, the artist is committing himself to a specific set of values (and depriving himself of valor). Regardless of which particular value system is adopted, the artist is the one who makes the choice between such value systems. Consequently, he cannot be relieved from the responsibility for his choice.

Therefore, the donkey's dilemma is a false dilemma: whichever principle he decides for, the artist will bear the burden of his choice. In either instance, the artist cannot detach himself from the moral and ethical implications of his decision regarding the artwork he creates and the guiding principle behind it. This is due to the fact that the artwork's constitution and composition are equal to a formal structure of co-organisation of various elements. By introducing an artwork to the audience, the artist presents it with a particular model of organisation and harmonization, such that it can be a model for the (organisation of) society. Thus, he bears moral responsibility for the created artwork and the model he provides – so much so that the dilemma is rendered false. In other words, there are different choices at the artist's disposal, hence the dilemma; however, the dilemma is revealed as false because only creation according to the inner principle escapes relativism. The artist is obliged to create a work of art following valor, or he (and his art) is to be valued as more or less useful to some frame of reference. The donkey is to be the one who carries Valor into a City or the one who carries others' (value) gold.

Bibliography

Beardslay, M. C. (2019), An Aesthetic Definition of Art. In P. Lamarque, S. Haugorn Olsen (eds.), *Aesthetics and Philosophy of Art. The Analytic Tradition; An Anthology*, Second Edition (pp. 22–29). Willey Blackwell.

Carrol, N. (1994), Identifying Art. In R. J. Yanal (ed.), *Institutions of Art. Reconsiderations of George Dickie's Philosophy* (pp. 3–38). Pennsylvania: The Pennsylvania State University Press.

Carroll, N. (2012), Essence, Expression, and History: Arthur Danto's Philosophy of Art. In M. Rollins (ed.), *Danto and His Critics*, Second Edition (pp. 118–145). Oxford: Willey-Blackwell.

Carroll, N. (2021), Essence, Expression, and History: Arthur Danto's Philosophy of Art. In Carroll, N., *Arthur Danto's Philosophy of Art: Essays*. Leiden/Boston: Brill.

Crowther, P. (2013), Indifferent to Intentions: The Autonomy of Artistic Meaning. In O. Hullat (ed.), *Aesthetic and Artistic Autonomy* (pp. 13–30). London: Bloomsbury.

Danto, A. C. (1973), Artworks and Real Things. *Theoria*, 39(1–3), 1–17.

Dillenberger, J. (2004), A Theology of Artistic Sensibilities: The Visual Arts and the Church. Oregon: Wipf&Stock.

Gaut, B., Livingston, P. (2003), The Creation of Art: Issues and Perspectives. In B. Gaut, P. Livingston (ed.), *The Creation of Art. New Essays in Philosophical Aesthetics* (pp. 1–32). Camebridge: Camebridge University Press.

Gover, K. E. (2018), *Art and Authority: Moral Rights and Meaning in Contemporary Visual Art*. Oxford: Oxford Unniversity Press.

Groys, B. (2013), The Logic of Equal Aesthetic Rights. In Groys, B., *Art Power* (13–22). MIT Press.

Grugan, A. (2009), Wassily Kandinsky: Creative Breakthrough to Absolute Art. In L. Della Pietra (ed.), *Perspectives on Creativity*, (pp. 2–25). Cambridge: Cambridge Scholars Publishing.

Hemingway, A. (2017), *Landscape between Ideology and the Aesthetic: Marxist Essays on British Art and Art Theory, 1750–1850*. Leiden/Boston: Brill.

Henry, M. (1973), *The Essence of Manifestation*. The Hague; Martinus Nijhoff.

McIver Lopes, D. (2014), *Beyond Art*. Oxford: Oxford University Press.

Lamarque, P. (2010), *Work and Object. Explorations in the Metaphysics of Art*. Oxford/New York: Oxford University Press.

Lamarque, P. (2013), Historical Embeddedness and Artistic Autonomy. In O. Hulatt (ed.), *Aesthetics and Artistic Autonomy* (pp. 51–63). London: Bloomsbury.

Osborne, P. (2000), *Philosophy in Cultural Theory*. London/New York: Routledge.

Roberts, J. (2023), *Art and Emancipation*. Leiden/Boston: Brill.

Thompson, J. M. (1999), *Twentieth Century Theories of Art*. Carleton University Press.

Šarović, S. (2021), 'Šta pada pod krušku? Reč o umetničkom metodu'. *Theoria*, 64(2), 161–171.

Young, J. O. (2001), *Art and Knowledge*. New York and London: Routledge.

https://doi.org/10.18485/baf.2025.2.ch9

Dušan Smiljanić

DISPOSITIONS – A PROPERTY OR A WAY OF BEING

Abstract: When we inquire about the nature of dispositions, we are asking, "Is a disposition some property or a way of being?" But, in asking that, we already assume the difference between a property as an ontic character and a way of being as an ontological character or entity. In both cases, a disposition is viewed as a characteristic of something, and so as something predicable, and not substantial. **Regardless, we now see that an adequate understanding of what a disposition is requires further investigation. In particular, it is necessary to examine whether the concept of ontological difference refers to something real. This concept involves the distinction between being itself—conceived as potentially manifold—and particular entities, such as properties or individual things. If ontological difference does indeed refer to something real, we must also ask whether it serves as a useful theoretical instrument in clarifying the nature of dispositions.** If it does refer to something real and is helpful, then our task would be to determine whether a disposition is a property or a way of being. If it is a way of being, then which way is it, and what is its inner structure? If it is not a way of being, then our task would be to determine which type of property it is and what it means to say that something is a property of property. Thus, our inquiry will be twofold: we will attempt to explore, by investigating the subject of dispositions, the question of ontological difference, and by investigating ontological difference, we will attempt to grasp the essence of dispositions.

Keywords: disposition, property, way of being, ontology, semantics, logic

1. Introduction: The concept of ontological difference and ontic-ontological distinction

The question of whether some phenomenon is a *way of being* or a *property* presupposes the notion of an *ontological difference*[1] as the con-

1 According to Heidegger, the ontological difference is the oldest, primordial difference, which precedes all science and philosophy, which is as old as language and humans alike. That difference is always implicit, even in our everyday language. When we

ceptual articulation of the distinction between being and entities. We distinguish entities as substances, then according to species and genera, properties and relations, but we distinguish being according to its *ways*. The sphere of entities and its divisions is called *the ontic sphere*. The sphere of being and its ways is called *the ontological sphere*.

The ontological difference is expressed negatively by the thesis that being is not another entity among entities. If every entity is an entity because it is something existent, and if the sum of all entities has something in common in virtue of which they are called entities, then that is a fact of their *being*[2]. Even though it is not some entity, being is not non-being; that is, it is not nothing. By saying so, we acknowledge that being, even if it is not a *something,* surely *somehow is.* The answer to the question of *how something is* has to be given through the exposition of the concrete *way of being of an entity*. *How something is* represents an interrogative form for *the structure of being*.

The conceptual articulation of the difference between the being (*Sein*) and entities (*Seiendes*) was taken up by Heidegger from his early days. Heidegger's term for this distinction is *Ontologische Differenz* or *Unterscheid*. According to Heidegger, far from being a product of our capacity to distinguish between what is and what is not the case, this differentiation of being and entities is the basis of that distinction. Philosophers "have traditionally presupposed this difference but failed to ask how it is possible (*GA* 14: 87n.; *GA* 70: 68; *GA* 71: 121, 126)" (Dahlstrom, 2021, 227). Or as Heidegger himself puts it: "The ontological difference is ... the difference ... in which being is distinguished from entities, the being that determines the latter at the same time in the constitution of its being ..." (*GA* 29/30: 521).

So, Heidegger not only acknowledges this difference as something real in the sense that it is not a mere notion without a reference in the world, but he postulates it as the foundation of his philosophical project.

Based on the ontological difference, Heidegger goes on to postulate a new distinction between *ontic* (*ontisch*) and *ontological* (*ontologisch*). The ontic concerns concrete properties and characteristics of an entity, in contrast to the ontological, which pertains to the specific way an entity of a certain kind has its characteristics. The early to mid-Heidegger centrally distinguishes between the ontic and the ontological, as his foundational

try to distinguish between two entities (*Seiendes*) A and B, and when we say "A is different than B", we already make an implicit difference between that "is" as a word for being (*Sein*) as well as A and B. For more on this, see in (Heidegger, 1982, p. 17–19).

2 Or we could say the fact that they *are*.

move is to keep entities (*Seiendes*) conceptually distinct from being (*Sein*). The adjective "ontological" concerns being – i.e., what it is for a given entity or class of entities to be – in contrast to the adjective "ontic," which applies to entities as such, i.e., their properties, their various arrangements and behaviors, whatever can be known empirically about them. When it comes to understanding the basic contours that delimit the possible ways for an entity of a certain kind to be, we are concerned with the ontological constitution of entities of this kind; when it comes to matters of this or that entity's concrete condition and properties, we're in the realm of the ontic. Thus, while physical objects, animals, and human beings are all entities, the ways in which they have characteristics – and accordingly the specific nature of their respective characteristics – differ fundamentally: physical objects are occurrent (*vorhanden*), animals live, *Dasein* exists. These terms indicate different ontological constitutions, each of which delineates a range of how entities of the respective types can be at all" (Slaby, 2021, 542).[3]

We do not need to go further in Heidegger's way of differentiating and articulating ways of being because we do not here want to see what Heidegger's concept of dispositions would be. Here, it is just important to stress the idea that when we are inquiring about the ontological status of some phenomenon, we have in mind the three possibilities: that the phenomenon can be an entity in a substantial sense, or some property, or some way of being of entity.

We are inquiring about dispositions here. Are they some being or a way of being? If they are some entity, it seems that it should be a property. If disposition is a property, what kind of property is it? If a disposition is not an entity but a way of being, what is its structure? How does a disposition determine the being of entities to which it belongs?

However, before we answer all these questions, we can wonder about the very manner in which we asked. **Is the concept of ontological difference ultimately justified, and if it is, what reasons can be given in support of its legitimacy**? What is this distinction for? Is the discourse about being and its ways even necessary, or is it reducible to the discourse about entities and their properties and relations?

---

3 Here, it is important not to identify Heidegger's ontic-ontological distinction with Peter Van Inwagen's distinction of the same name since for him, that difference concerns the division of arguments. Inwagen calls ontic that infinite class of valid arguments in which one starts from a premise that sets some set of properties as a sum of certain conditions, which, if fulfilled, in the conclusion, it turns out that there is something that exemplifies those properties. On the other hand, he calls ontological arguments that start from the same premises ontic, but in the conclusion, the exemplifying entity turns out to be God (Van Inwagen, 1977, 375).

Therefore, the purpose of this inquiry is twofold. On the one hand, concerning the example of dispositions, we want to examine the justification and appropriateness of the concept of ontological difference. On the other hand, the examination of the ontological difference and eventual legitimization of the discourse on the ways of being should help us illuminate the nature of dispositions.

2. The grounding puzzle of the question of dispositions

What does it mean to say that something is dispositional? **In the philosophical literature, dispositions are typically understood to be properties. They are either regarded as a specific kind of property or as expressing the general nature common to all properties. Alternatively, some accounts interpret dispositions not merely as properties, but as a distinct way of being.**

This means that, on the assumption of the ontological difference, the question of the nature of dispositions is the question of its ontic or ontological character.

But actually, the real puzzle in the question of the nature of dispositions lies in its ontological status. By this, we mean that because nobody would deny the existence of the phenomena that are recognized as realizations or manifestations of dispositions, the difficulty becomes the task of articulation of the being of the dispositions when they are not yet realized. So, if we accept that dispositions *somehow are* (that they are real) when they are not realized, then the question is whether they are properties, way of being of properties, or ways of being of entities themselves.

Some suggest that dispositions, when they are not realized, are nothing real, but simply our way of talking about entities and their future events. That is the grounding puzzle.

The same problem was at the center of the Antic discussion about the ontological status of δύναμις. Let's take a brief look at that subject.

3. Origin of the concept: Aristotle's articulation of δύναμις against Megarians reductionism

We owe the concept of a disposition, like many others, to the very beginnings of ontological thinking. We find the first clear articulation of this notion in Aristotle's concept of δύναμις, which was originally understood in unity with ἐνέργεια and/or ἐντελέχεια and according to movement (κίνησις). Before Aristotle, δύναμις meant force, power, strength, possibility, ability. In Aristotle's interpretation, δύναμις becomes one of the fundamental ontological concepts used to name the specific way of being of almost all entities in the world[4]. For Aristotle, therefore, what we call a disposition is undoubtedly the way of being [5].

With this concept of δύναμις Aristotle tried to avoid, on the one hand, Parmenides' static ontology[6] and to defend the phenomenon of *movement* (in all its modes of "emergence, decay, multiplication, diminution, transformation, change according to place" (*Cat*. 15a13)) from banishment to bare appearance and illusion, and on the other hand to deny the Megarian thesis of the identity of δύναμις and ἐνέργεια by reduction to the absurd consequences[7]. By this, Aristotle wanted to show that between being and non-being, *there are* (manifoldness – πολλά/πολλαχῶς) *ways* (τρόπον/τρόποι) *of being* (εἶναι) that enable the transition from one to the other through movement. He listed the most fundamental ways of being in the scheme of categories.[8] Apart from categories, Aristotle dealt with other ways of being in almost all other writings, and among the fundamental ones he placed four: being in the sense of the scheme of categories (κατὰ τὰ σχηματα τῆς κατογοριας), being in the sense of δύναμις and ἐνέργεια, being in the sense of truth and lies (ὡς ἀληθὲς ἤ ψεῦδος), and being as an accident (κατὰ συμβηβεκός)[9]. Aristotle's understanding of δύναμις as a way of being will also be illustrated by the example of the fragility of a glass.

---

4 Aristotle explicitly writes (*Meta*. Θ, 1046a8): "καὶ δυνατὰ καὶ ἀδύνατα λέγομεν τῶ εἶναι πως ἤ εἶναι" – "We also speak of this way of 'powerful', and 'powerless' to the extent that something is or is not in a certain manner".

5 One author who agrees with our reading of Aristotle, but only with different terminology, is Witt: "... Witt (2003) argues that Θ investigates potentiality and actuality as ways of being, which may apply to any of the kinds of beings" (Louise Gill, 2005: p. 30); and also: "According to Witt, Q contains two main arguments, first concerning the existence of δύναμις (Θ, 3), which she thinks can mean either causal power or potentiality (= inactive power), and second concerning the priority of actuality or activity in relation to δύναμις (Θ, 8)" ... (Louise Gill, 2005: p. 64).

6 For a classic interpretation of Parmenides ontology see in: (Anscombe, 1981, p. 3–9), (Arsenijević, 2007, p. 133–140).

7 More on this topic, precisely on the so-called Megarians and their teaching, see in: Chame, S. (2024), On the Megarians of *Metaphysics* IX 3. *Archiv für Geschichte der Philosophie*, 106 (2), 177–206.

8 I showed that in the paper *Aristotle's teaching on categories in light of question of the relation of language and reality* presented on student conference in Belgrade in 2024 (text is still in print), that the scheme of categories has its own ontological foundation and that the categories have their ontological and even ontic sense in addition to their linguistic-logical meaning.

9 The question about the prior unity of these four senses of being was the problem of special attention in Aristotle, but for answering it, we have no time now.

Namely, when we say that a glass is fragile, we mean that it can break. It has fragility as its internal possibility. Fragility is part of its δύναμις. However, while the glass is fragile, it means that it is not broken. How is it that when something does not yet exist, we can attribute it positively? This upset Parmenides and Megarians. Parmenides declared it as a *contradictio in adiecto* and therefore dismissed it as meaningless discourse. The Megarians denied the difference and, through reduction, identified the dispositions with their actuality (ενέργεια). That is, according to Parmenides the glass is not breakable, because breaking as such is not possible, because it represents a kind of process or change that implies movement, and movement is the transition from being to non-being, and any discourse about non-being is meaningless and false, for the Megarians the glass is breakable, but only when and while it breaks (*Meta*. 1046b29–30). There is no such thing as a fragility that precedes the breaking itself. Fragility and breaking differ linguistically, *but only linguistically*, and this difference tempts us to attribute it to reality itself. For the Megarians, the discourse on fragility before and independently of actual breaking does not refer to anything actual, so is, strictly speaking, either unnecessary and meaningless or reducible to the discourse on actual breaking. Aristotle rejects both solutions.

As a principle element of his philosophy, Aristotle accepted the thesis of "saving phenomena"[10] and the so-called "common-sense intuitions of the sacred" whenever and as much as possible. Since the discourse on dispositions is implicit in everyday life, which takes place in parallel and independently of any philosophical and scientific theory, according to Aristotle, it should be preserved if possible. We mark the difference between movements and changes and breaking and fragility in everyday life. So, one must see what the difference is if he wants to neglect the Parmenides-Megarian thesis.

3.1. Aristotle's positive solution

To reject a Megarian[11] identification of, or reduction of δύναμις to actuality (ἐνέργεια), Aristotle provides a few arguments with the same logical structure. They are constructed in such a way that from the (Megarians) premise that δύναμις is identical with actuality, and that when there

10 This refers to the self-evidency and reality of the ways in which thing looks (appears) to us before any theoretical considerations.

11 "The traditional view of the Megarian school takes Diogenes's testimony at face value; that is, it holds that the school's doctrine resulted from a synthesis between Socratism and Eleaticism" (Chame, 2024, 3).

is no actuality, there is also no such a thing as δύναμις, we come to the absurd consequences.

Aristotle starts from the example of learnable skills, which are acquired dispositions (δύναμις) (1046b30–1047a5). In the case of painting, if we accept that there is no skill of painting in a painter, or more precisely that a painter does not possess it when he does not paint, then we are forced to conclude that whenever a painter starts to paint again, he in some way re-acquires the skill. Because that "mysterious" re-acquairing of skill appears absurd, Aristotle concludes that the premise is absurd. This is the case for all learnable skills.

The second argument is about the dispositions of non-living entities, and the third considers natural dispositions of living beings (1047a5–1047a10). The second argument is more complex because it refers to Aristotle's theory of sensual perception and denies Protagoras' theory of perception. This argument could be considered circular because already in his work on the soul *De Anima*, Aristotle uses the distinction of δύναμις-ἐνέργεια as grounds for neglecting Protagoras' teaching. So, we can't go deep here about that.

In the case of the natural dispositions of living beings, the argument is structurally analogous to the first one. If we accept that when some man or animal does not possess the faculty of perceiving, we are then forced to conclude that a sleeping man is blind and deaf and that he immediately acquires that faculty when he gets up. This goes for all dispositions that are natural as a type of δύναμις, such as fertility, physical strength, the capacity for speaking, walking, etc.

The last argument is modal, and it is about postulating a real (ontological, not just lingquistical or conceptual) difference between δύναμις and ἐνέργεια, and to again reduce the absurd the thesis about the identity δύναμις and ἐνέργεια (1047a10–1047a20). **For if possibility is identical with actuality, then we must deny that anything which is not actual can exist in any way whatsoever**. By that, we are left only with the distinction of actual and impossible. As Aristotle suggests, this thesis implies that motion and generation are impossible because that which is standing will always stand, and that which is sitting will always sit; because if it is sitting it will not get up. Because of Aristotle's principle of the preserving phenomena and the self-evidence of being of motion, change, and generation in experience, Aristotle concludes δύναμις and ἐνέργεια are different.

With these arguments, Aristotle succeeded in making space for accepting the necessity of differentiating δύναμις from ἐνέργεια. But we still do not see clearly how δύναμις should be understood, whether as some property, type of properties, some entity in a substantial sense, or as a way of being.

But one thing is certain: The Crucial point about this debate is not Aristotle's counter-argument, but that debate is still alive, and continues on the same grounds. A *Megarian thesis is implicit in all modern reductionist views on the nature of dispositions* such as Carnap's or Ryles's (1949) views. And so, if it can be decided whether Aristotle can stand up against it, then neglecting modern versions of reductionism may have to be grounded on Aristotle's concept of δύναμις.

We believe it is so. Thus, we will first show in which aspect Aristotle's concept of δύναμις is better and different from the modern dispositionalist view, and then in what way it can be made stronger.

So, first of all, let's take a brief look at two basic contemporary views on dispositions. One, which is in line with the Megarian thesis in some sense, says that dispositions in a non-actualized state are nothing more than a special type of statement – a discursive category. Then, the other view, on which dispositions are taken to exist before their realizations, but as some property.

4. Contemporary debate

To the question of dispositions in the contemporary discussion have come two groups of possible answers[12]. For the first group, which I call *the linguistic view*[13] on dispositions, dispositional terms are just linguistic categories. It's a way of talking – a special type of ascription about objects, their properties, and future events[14]. For them, there is not anything in the world to which the predicate "dispositional" refers, besides objects and their behavior. For the other group, which I call *the property view*, dispo-

12 Clark has a similar view: "And the first thing to point out about the contemporary distinction is that there is not just one, but in fact (at least) two distinctions that use the language of 'categorical' versus 'dispositional'; and each of those two distinctions has a different subject matter. One of them does not have any direct implications on metaphysics but rather is a distinction between two sorts of ascription or two sorts of truths. The second distinction concerns the nature of properties, and thus does purport to be straightforwardly metaphysical" (Clark, 2014).

13 This idea can be traced back to the early 20th century, when philosophers from the Wien circle had a construction of a totally "empiricist language in which all meaningful sentences including dispositional sentences can be analyzed in observational and extensional-logical terms" (Fara & Choi, 2018).

14 For example, Carnap or Ryle (1949).

sitional terms do have reference, and they refer to actual properties possessed by objects[15].

Of course, these main two groups are very divided in themselves by questions such as what kind of ascriptions are sufficient for speaking about dispositions or whether all properties are dispositional versus just some of them, etc.

I will be defending the thesis that claims the following things about the nature of dispositions:

1) *The discourse about dispositions does not consist just in special types of ascriptions about future events, but it has reference even when dispositions are not yet realized.*
2) *Dispositions are not properties at all.*
3) *Dispositionality consists of a way of being of entities.*

4.1. Dispositions are not just a type of ascriptions

I have two major objections to *the linguistic view* on dispositions:

a) The purpose for this kind of reduction of dispositions to ascriptions of the future events and possible behaviors of objects lies in the quest to deny the thesis that dispositions are some existing properties, even when they are not realized. So, the problem with this strategy is that reduction is based on the assumption that dispositions are properties, which is, in my opinion, wrong. The whole strength of the *linguisticist* argument lies in the weakness of the *property view*. Arguments for this thesis will be given in the next **section**.
b) As I already mentioned, *the linguistic view* is just a more sophisticated version of the Megarian thesis. Even if it is a more sophisticated version, I still mean that Aristotelian contra-arguments hold for them in the same sense as for the Megarians because Aristotle didn't hold the property view on dispositions, but he considered δύναμις a way of being.

15 "So *to be categorical* is *to be actual* or *occurrent* —"here and now." Dispositionality, by contrast, is a matter of *what things do owing to their circumstances*. Heil's view of properties differs from Lowe in that, for Heil, disppositionality is, indeed, "built into the properties." That is to say that properties themselves are irreducibly dispositional... So, like Lowe, Heil understands the categorical—dispositional distinction to *reside at the level of predication*. He states that: "A property's dispositionality and its qualitativity are ... the selfsame property differently considered." ... For Heil there is nothing beyond the properties' natures themselves that make dispositional ascriptions true" (Clark, 2014).

4.2. Dispositions are not properties

The main problems with any *property view* on the nature of dispositions are these:

a) It does not recognize the (ontological) difference between a way of being and a property.
b) It does not possess a clear concept of property.
c) It has trouble explaining the being of dispositional properties when they are not actualized, which is needed to avoid reduction to categorial properties.

These three theses are connected such that the second and the third follow from the first one. It all starts with ignoring ontological difference and the ontic-ontological distinction. It's not important here to consider or even accept a Heideggerian way of distinguishing ways of being. It is sufficient to accept this difference itself as the starting point of inquiry, and then to argue that the question of dispositions lies in the ontological sphere. That is, it is the question of the ontological status of entities, their way of being, and not of the ontic sphere, which considers the properties and other ontic characteristics of entities in a way thate science approaches it.

So, from the assumption of ontological difference, the identification of a property with a way of being is possible only in the numerical, but not structural, sense. By that we mean, because every entity has its way of being, every property has a way of being. And in a numerical sense, we can't differentiate between a property and its way of being like we differentiate one property or entity from another. But that does not mean that the property is the same "thing" as its way of being. The reason for this is two-fold. First, all ontic characteristics of entities, with entities themselves, are somehow grounded in their ways of being. And second, from the ontic point of view, besides properties, we can't find anything else in entities, but properties are not all characteristics that could be found in the entities because there is another ontological point of view.

So, if we want to prove that dispositions are not properties, we only have to look at entities to which we ascribe dispositions from the ontic point of view and to look to where dispositions are. Briefly, we will find nothing. Look at the glass, for example, when it is not broken. Where is its fragility? Or look at any of the five senses of a man when he is sleeping. Where are perceptions? Or take a look at the man who is good at football whether he is lying in bed or playing the game. Where is that ability, in his feet or in his head?

But if we can't find these dispositional "properties" beside the "categorial" properties of entities, we can say that dispositional properties are the same properties that some call categorial. From this step, we can conclude, as some *property view* defenders do, that all properties are dispositional and that their dispositionality consists in its casual powers, or some inner nature, or ability to manifest itself under certain circumstances. But in this we didn't achieve anything. Because we again fall on *the linguistic view* which stresses discourse on future events and conditions for manifestations, with the difference that dispositional properties do exist as characteristics of entities independently of their manifestations, but their ontological status – the way in which dispositonal properties are – remain in "fog".

This step also reveals that for the clear concept of property, whether "categorial" or "dispositional", it's necessary to have already articulated a way of being in virtue of which something is a property and not a relation or particular entity. And the importance of accepting the ontological difference becomes more obvious.

So, the main task here is to escape reduction of any sort. Both *linguistic* and *property views* involve reduction. The *linguistic view* is more explicit: they openly say that there is no such thing as a disposition without a realization, there are just things in the world that can do something under certain conditions, and some that cannot. This means dispositional ascriptions have their reference only when comes to realization, i.e., in the future.

Defenders of the *property view*, because they operate with only ontic view on dispositions, actually do the same thing because they connect dispositions with realizations, with the only difference being that they claim that dispositional terms have reference before realization. But that reference is "some mysterious inner essence" that they call dispitional, or it is about non-claimed reduction to the actual properties or states of affairs of objects.

From this it follows that the only solution for escaping reduction of any sort is to postulate dispositions as a way of being, which somehow are present in the ontological structure of entities before realization. The reason for that option is the same as Aristotle's, which we already mentioned.

5. Dispositions as a way of being

And now we come to the crucial question of this work: Is a disposition a way of being of entities, or properties, or relations, or all three of them? But, before this, let's briefly elucidate the concept of a way of being.

5.1. Aristotle's criteria

The first, but not the most clear and sufficient articulation of the concept of a way of being, as far as I know, is found in Aristotle's work. Often, in his text, we find the question of *the way* or *the manner in which* some phenomenon *is*. For Aristotle, that was one of the most basic and important philosophical questions. In his probably first or earliest work, *Categories*, he implicitly used what I call *ontological criteria* for differentiating things under the different categories. *That criteria consist of revealing the way of being of some phenomenon, by which that thing can be put under some of the categories*. **Expressed in reverse order**, every category, as part of its concept, has conditions that the entity has to fulfill to get subsumed under it[16]. We are now focusing on the ontological.

Take, for example, the two categories, the first category of substance (οὐσία), and the second, of relation (τὰ πρός τί). Things from the world that can be named as substances have to fulfill these ontological criteria; i.e., they have to be in this way:

- The thing has to be only in virtue of itself, that is, to have independent existence
- The thing has to be as one and nondivisible in its being
- The thing in virtue of its being has to be capable of possessing other things which can change, without losing its identity, etc..

---

16 I think that we can find three senses of the category:

1. *The grammatical sense* of categories: Categories are uncomposed linguistic expressions, from which sentences are created by assembling, like any claims about something.
2. *The logical sense* of categories: Categories are those unassembled linguistic expressions, by combining which judgments are formed. Bringing about judgments about being it is determined (κατηγορεῖν) in a potentially true or false way. Categories outside judgments have no truth value, and in judgments they can occupy two possible logical functions, of subject or predicate.
3. *The ontological sense* of the categories is divided into two parts, hierarchically arranged as follows:
   α) *The purely ontological sense* of the category consists in the presentation of way of being of the thing, grounded in two basic ways of being as subject or as predicate.
   β) *The ontic sense* of categories consists in naming and thus referring to entities and their characteristics.

What has puzzled interpreters for centuries is that Aristotle names linguistic expressions with categories that refer to reality and the logical functions of those expressions and the very aspects of reality that are named by the expressions, without clearly explaining in the text which of these meanings the category signifies, but it must be inferred from the context of occurrence.

So, for a thing to be called a substance, a way of being of an entity has to be constituted like this.

Also, for the category of relations, conditions are such that things are something only in relation to something else. That being only in relation to its other way of being (relational) from the substantial one.

And we agree with Aristotle that a disposition or δύναμις is a term that refers to some specific way of being that does not fall under any of the categories, but is present in every category of being, i.e., in some sense, present in almost everything. Now we have come to the point when we have to describe this peculiar dispositional way of being.

What Aristotle failed to recognize — regardless of whether he was on the right path in his investigation — was precisely what Heidegger succeeded in discerning. Heidegger recognizes that every way of being has its constitution or structure. But also, there is one big and basic difference between the human way of being and the ways of being of all other entities in the world.

5.2. Heideggerian concept of the structure of the way of being

For Heidegger, every way of being (*Sein*) of entities (*Seiendes*) consists of some structure. Because of that, "Heidegger's approach to ontology is structural in the sense that he generally proceeds by identifying the structural features constitutive of the being of entities ..." (Wrathall, 2021, 705). As Mark A. Wrathall rightly continuous: "A structure is an organization of constituent features of a whole – an organization in virtue of which that whole can perform a function. Thus, Heidegger explains, structure is "read off" of a function (*GA* 1:417/BH 62). "One formally describes a structure by specifying the types of relationship that necessarily must obtain between types of elements in order for the whole to perform its function. Thus, everything that performs a certain function will possess those types of elements standing in that type of relation **...** The structure opens up and constrains the function that the entity can perform" (Wrathall, 2021, 705).

We do not here go deep into Heidegger's ontology and his differentiation of different ways of being. Most importantly, we must make explicit the basic differences between Heidegger's and Aristotle's articulation of the notion of the way of being (i.e., actually the whole metaphysical tradition between the two mentioned thinkers).

The basic differences are:

a) *Absolute difference between human and non-human*[17] *ways of being*

For Heidegger, *existence* (*Existenz*) is a term that denotes only the human way of being. That means that only humans *exist*, while all other entities *are* in some other ways. Heidegger's technical term for human entities is *Dasein*. Literally, this German word means *being-there*. Before Heidegger, the term *Dasein* had almost the same meaning as the terms *Existenz* (existence) or *Sein* (being), and all three of them were used for all entities without any difference.

The basic difference between humans and all other entities is that despite the fact that all entities somehow are, only humans are "conscious" of their being and being in general. That is, only humans possess some understanding (*Verstehen*) of being (*Sein*). This fact is a ground point for human possibility, not just in philosophy, but for being human at all. That's why humans are called *Dasein* or being-there, because they are constituted in such a way that we are "the place" on which being itself comes to the word, the appearance, revelation, and understanding of itself.

So, human beings possess their own structure of their way of being. That structure is called ontological-existential because existence is the name for humans' way of being. This leads us to the second basic difference:

b) Ontological and epistemic primacy of articulation of the human way of being over the conceptual articulation of the ways of being of non-human entities and the concept of being in general

In the first chapters of *Being and Time,* Heidegger gives an argument on why it is necessary, in order to get a full understanding of being in general, to start from analyzing one peculiar entity and its way of being. That entity is, of course, *Dasein,* and its way of being is called *existence*. **The idea is the following: due to the inseparability of any way of being from the concrete entity that is in that way, it is necessary, for the sake of obtaining the concept of being in general, to first determine which**

17 Here it is important to note one thing about the "human" and "non-human" way of being distinction. Namely, Heidegger consciously and deliberately refrains from using the term "human" when he talks about the beings that we ourselves always are, but instead uses the already mentioned technical term *Dasein*. He thereby wants to avoid the traditional metaphysical assumptions and implications with which the term "human" is saturated. Another reason for such a decision is the possibility, which Heidegger shows, that human is spoken of in all three senses (or ways) of being, which are *presence-at-hand* (*Vorhandenheit*), *readiness-to-hand* (*Zuhandenheit*) and *existentiality* (*Existenzialität*), while Dasein can only be attributed to existentiality. However, in order to aide the potential reader, we will not stick strictly to Heidegger's terminological guidelines, and will retain the discourse on "man".

entity can serve as the exemplar for acquiring this concept. Reasons for depicting *Desein* and analyzing its existence are as follows:

- Dasein **is a unique entity in that it alone is capable of raising the question of being**.
- This possibility lies in the core of the structure of *Daseins*' way of being (existence)
- **The structure of Dasein's way of being is most accessible to Dasein itself.**

This is why for Heidegger, in place of "first philosophy" should be what he called *fundamental ontology* or *existential analytics of Dasein*, and not the conceptual articulation of being as such (*being qua being*) as it is in Aristotle.

These things cannot be found in Aristotle's ontology. For Aristotle, the difference between human and non-human entities is not philosophically that fundamental. The way of being of humans is not radically different from the ways of being of other entities; in fact, for Aristotle, humans exist as a special type of living inanimate entity who have some higher properties then "regular" animals.

These two assumptions have special implications for comprehending the nature of dispositions, which we see as a further development of the Aristotelian ground. We will make this clear in the final part of the paper.

Keeping this in mind, we will first try to prove the necessity of comprehending dispositions as ways of being of entities and not as some property or discursive category. Then, in the final **section** we will see how we can get to the final articulation of the concept of a dispisition as a way of being through a kind of synthesis of Aristotle's and Heidegger's ontological achievements.

6. Three examples to prove dispositions as ways of being

There are three basic types of dispositions, and by analyzing exemplary cases of each type, we will show that dispositionality consists of a special way of being of an entity and that it is not a matter of properties or a mere linguistic category. The basic types of dispositions are natural or innate dispositions of living beings, acquired dispositions of human beings, and "natural" dispositions of inanimate beings such as natural elements or objects of the human world.

We will take examples of the three basic types of dispositions: the *disposition of motion* as a natural disposition of living beings, *reading ability* as an acquired ability of humans, and *freezeability* as the "natural" disposition of water as one nonliving entity.

6.1. Disposition of motion

One of the fundamental dispositions of all living beings is the capacity for movement in the sense of change of place. Let's consider the disposition to move to be a mere linguistic ascription about the future that speaks of how living beings can behave under certain circumstances, without saying anything positive about them when they are at rest. This can be answered as Aristotle did, by noting that it follows from there that if the disposition to move does not exist outside of movement itself, then living beings actually and "mysteriously" acquire this disposition every time they move.

But does it follow that the disposition of motion is some kind of property? We will easily see that it is not. Because nowhere on any part of the body of living beings do we find the disposition of movement among other ontic characteristics such as size, shape, color, arrangement of organs, etc.

The disposition of movement must then be the way of being of living beings. They are the ones that can move or be at rest. The *being* of living beings, as inseparable from the entities themselves, is that which passes from the mode of rest to the activation of its innate disposition of movement and *vice versa*. Movement as a function is a structural moment of the way in which living beings are.

6.2. Reading ability

People have the disposition to read. It is about the ability to interpret the sign system. It is unnecessary to give a precise and complete definition of the phenomenon of reading here. We are interested only in the ontological status of this disposition when it is not realized, i.e., when a person does not read either because he is sleeping or doing something else while awake. Is reading, taken as a paradigmatic example of an acquired human disposition, at all something real beyond its realization, and if so, is it some kind of property or a way of human beings?

Let's again consider the ability to read as a linguistic ascription of the future that speaks to how a person can behave under certain circumstances, without saying anything positive about a person when he does not read. This can also be answered as Aristotle did, by noting that it follows that if the ability to read does not exist outside of reading itself, then every time a person starts to read, he suddenly learns or "mysteriously" remembers or learn these skills.

That is, of course, absurd. How, then, is reading then a property? It seems that it is not, because if we look at the human body, and its organs used for reading – eyes, brain, and hands – we will not come across the ability to read anywhere. If it is not even a property, therefore, something ontic, then it seems must be a way of being.

The disposition to read is, therefore, one way in which a previously literate person exists. This disposition is the name of an acquired function, through the acquisition of which man enriched the structure of his existence, with an additional possibility of his existence. And indeed, it is absurd to attribute the disposition of reading to any part of man, instead of to man as a whole individual being. Human existence is not something material and ontic, and its structural moments, functions, and dispositions cannot be located among its ontic and material characteristics.

The way in which a disposition exists when it is not active is *possession*. By moving into actuality, changes occur again in the way of existence. That is, the same human existence changes from passive possession to active disposal of that power when a person gets hold of a text. The very existence of an entity that is the subject of an acquired disposition is *dispositional*.

6.3. Freezeability of water

As for the dispositions of inanimate entities, let's take, for example, the disposition of water to freeze at a certain temperature. For the same Aristotelian reasons, we see that it is absurd to claim that the freezeability is merely an ascription of possible future responses to temperature. Because that would mean that the water is freezable only when it freezes. This is not the case because when it's cold, water freezes, and it freezes in the cold because as such, and therefore by itself, it freezes in cold conditions.

Also, it is obvious that we cannot find such a thing as freezeability among the properties of water when it is in a liquid or gaseous state. Each aggregate state of water is a way of its being, i.e., a set of functions of possible behavior under certain conditions. We see that the disposition to freeze, therefore, is the way in which water is in a liquid state.

7. Conclusion – Final articulation of the concept of a disposition (on the relationship between Aristotle's and Heideggers approach)

So, if it is rightly justified that disposition should be seen as a way of being, we now have to see where the differences are between Aristotle's and Heidegger's concepta of a disposition as a way of being. And most important, we will ask: are these differences in contradiction or can they be synthesized in a wider picture?

We will start from the common points on which Aristotle and Heidegger are in agreement:

- Dispositional terms denote the various ways of beings of entities and not some properties or just linguistic categories
- The dispositional way of being is one of the fundamental ways of being common to almost all entities
- Dispositions are something "real", that is, something which is grounded in the entities themselves even before its realization and manifestation
- There are various types of dispositions

This list could surely be longer, but this is sufficient for us now. So, let's take stock of the differences.

- Aristotle does not mark the difference between human and non-human ways of being in the way that Heidegger does, and because of that, he does not make that radical distinction between humans' and non-humans' dispositions.
- While for Aristotle ἐνέργεια have ontological, temporal, and logical priority over the δύναμις in the case of dispositions, **Heidegger does not accept this primacy**. **For Heidegger, things are opposite in that for** Daseins, **dispositions have ontological primacy over their actualization.**
- This is because of human freedom. By cotrast, according to non-human entities, Dasein does not have priviously determinated purposes of its existence, but has the freedom to choose and even create it on his own. Heidegger says: "In each case Dasein *is* its possibility, and it 'has' this possibility, but not just as a property [*eigenschaftlich*], as something present-at-hand would. And because Dasein is in each case essentially its own possibility, it *can,* in its very Being, "choose" itself and win itself; it can also lose itself and never win itself; or only "seem" to do so" (Heidegger, 1962, 68).

Is this a problem for a task of synthesis of this sort? Can Aristotle's and Heidegger's notions on dispositions be made coherent?

In my opinion it can be done.

Although they are using different terminology, Heidegger and Aristotle reach many common results via different paths. Nevertheless, on the example of dispositions, there were manifestations of those basic differences in approaches between the two thinkers. Aristotle could not distinguish between dispositions as Heidegger does because the ontological difference was not laid down and illuminated to the extent that it was for Heidegger. Then, in terms of understanding the nature of man and the difference between humans and non-human beings, Aristotle did not see such a gap as Heidegger posited.

That is, the teleology that Aristotle discovered on objects of human production, he used as an explanatory ontological model for the whole of the cosmos, and thus in return for human nature itself, and this is what Heidegger tries to avoid. For him, such extensions are an example of "forgetting of being" because in these cases there is always a tacit and often unconscious identification of human and non-human ways of being. That is, the absolute difference between them is forgotten, in order to move from the difference in "*esse*" to the differences in terms of "*res*", i.e. properties, which represents a fall into the ontic in science and a fall into the *reification* of human nature in the ontological plane.

But by accepting Heidegger's insistence on the primacy of possibility over actuality in the case of human dispositions, the scope of Aristotle's principle of the primacy of ἐνέργεια over δύναμις is just narrowed to essentially nothing else. This narrowing is justified on the assumption of prior acceptance of radicality in the difference between human existence and the ways in which other non-human entities exist. Acceptance of this difference can also be understood as a supplement and a deeper and further elaboration of Aristotle's attempt to demarcate humans and non-humans by means of λόγος, νοῦς, political association, etc.

In this sense, I think that concerning the ontological status of dispositions, after Aristotle and Heidegger, there is nothing more significant to say, which does not mean that there is nothing more significant to say and discover about dispositions in general. I believe that the approach to dispositions must be focused on research concerning concrete phenomena, primarily in the human world of spirituality, culture, politics, etc., which after an ontological examination were revealed as dispositional. Through such analysis, we pave the way for further scientific investigations and better understanding of the thematized phenomena.

References

Aristotel, (1992). *Kategorije*. Zagreb: Medicinska naklada.

Aristotle, (1999). *Metaphysics*. London: Penguin Classic.

Arsenijević, M. (2007). *Prostor, vreme, Zenon*. Novi Sad: Izdavačka knjižarnica Svetozara Stojanovića

Anscombe, G. E. M. (1981). *The Collected Philosophical Papers of G. E. M. Ansombe: Volume One: From Parmenides to Wittgenstein*. Oxford: Basic Blackwell.

Clark, E. D. (2014). "How Aristotelian is Contemporary Dispositionalist Metaphysics? A Tale of Two Distinctions", *Proceedings of the American Catholic Philosophical Association*, 88: 85–99

Chame, S. (2024). "On the Megarians of *Metaphysics* IX 3", *Archiv für Geschichte der Philosophie*, 106 (2), 177–206.

Dahlstrom, M. A. (2021). Difference (*Unterscheidung*). In Dahlstrom's, M. A. (eds.), *The Cambridge Heidegger Lexicon* (pp. 227–231). London: Oxford University Press.

Inwagen, P. V. (1977). "Ontological Arguments", *Noûs*, 11(4): 375–395

Heidegger, M. (1978). *Frühe Schriften: (1912–1916)*, 2nd edn., ed. Von Herrmann Friedrich-Wilhelm, Klostermann, Frankfurt am Main.

Heidegger, M. (2007). *Zur Sache des Denkens: (1962–1964)*, ed. Von Herrmann Friedrich-Wilhelm, Klostermann, Frankfurt am Main.

Heidegger, M. (1983). *Die Grundbegriffe der Metaphysik: Welt – Endlichkeit – Einsamkeit (1929–1930)*, Klostermann, Frankfurt am Main.

Heidegger, M. (2005). *Über den Anfang*, ed. Paola-Ludovika Coriando, Klostermann,

Heidegger, M. (2009). *Das Ereignis (1941/42)*, ed. von Herrmann, F.W., Klostermann, Frankfurt am Main.

Heidegger, M. (1962). *Being and Time*. Oxford: Blackwell Publishers.

Louise, M. G. (2005). "Aristotle's Metaphysics Reconsidered", *Journal of the History of Philosophy,* 43, 223–51.

Witt, C. (2003). *Ways of Being: Potentiality and Actuality in Aristotle's Metaphysics*. Ithaca, N.Y.: Cornell University Press.

Ryle, G. (1949). *The Concept of Mind*. London: Oxford University Press

Smith, Ch. P. (1986). Translator's introduction. In Smith, C. P. Smith (eds.), *The Idea of the Good in Platonic-Aristotelian Philosphy* (pp. vii-xxxi). New Haven and London: Yale University Press.

Slaby, J. (2021). "Ontic (*ontisch*)". In Dahlstrom's, M. A. (eds.), *The Cambridge Heidegger Lexicon* (pp. 542–547). London: Oxford University Press.

BALKAN
ANALYTIC
FORUM

BAF2+: Dispositions
and Values

5. DISPOSITIONS AND VIRTUES

https://doi.org/10.18485/baf.2025.2.ch10

Amber Riaz

MORAL LEARNING FOR THE WRETCHED OF THE EARTH

Abstract: According to the most plausible contemporary accounts, moral learning involves the "affective attunement of reasons" through experience. The reliability of our moral intuitions is partly a function of this attunement and the environment in which it occurs. However, this account is incomplete regarding the moral learning of epistemically non-ideal adult agents who were raised in defective moral environments. Once immersed in healthy environments, such agents develop conflicting intuitions and a sense that their original intuitions have a faulty track record. This internal conflict and epistemic self-doubt constitute a kind of moral learning for such agents; they also play a significant role in developing moral understanding. "Wide reflective equilibrium" and its cousin, "consistency reasoning," are of limited to no use for them.

Key words: Moral intuitions; concept possession; affective attunement; dual process; reinforcement learning; coherence; consistency; fragmented mind

1. Introduction

Consider the following two cases.

– Hassan was raised in a morally healthy environment with zero tolerance for social evils like sexism, racism, homophobia, and discrimination against minorities or other species. Consequently, it is no surprise that he navigates the social world with ease. Hassan intuitively and reliably differentiates between morally appropriate and inappropriate responses. Although occasionally out of his depth in unfamiliar situations, he quickly develops an accurate understanding of the new social dynamic.

– Conversely, Abdul was raised in a toxic moral environment that thrived on patriarchy, misogyny, and other social ills. As a result, his intuitive responses to moral situations are as unreliable as Hassan's are dependable. For instance, when Abdul judges an act as right because it is a sign of respect and he might even cite some relevant moral principles to explain his judgement, it is often, in fact, an act of oppression.

Question 1. What does Hassan have that Abdul lacks?

Answer. Both Hassan and Abdul share some key thin and thick moral concepts, such as "right", "wrong", "respect etc., and abstract moral principles. However, Hassan is skillful in applying them, while Abdul is not.

Question 2. Can Abdul become like Hassan if he is immersed in Hassan's environment as an adult?

Answer. It is unlikely. Hassan's moral development is the result of a gradual affective attunement to moral reasons from an early age. This process, on its own, is unlikely to be of much benefit to Abdul's moral learning as an adult.

Question 3. What are Abdul's options as an epistemically non-ideal and morally fallible agent striving for moral improvement and learning?

Answer. His primary option is reflection, but not necessarily aiming for a reflective equilibrium or simple consistency in reasoning.

The following sections explain these questions and answers.

2. Concept Possession

The idea that both Hassan and Abdul share at least some of the same moral concepts is not philosophically innocent. It assumes a specific theory of concept possession, the JAZZ model of meaning on which a speaker must satisfy two conditions to be competent with the meaning of an evaluative term. "*Coordinating Intentions*: The individual speaker must have a coordinating intention to use the term in a way that makes best sense of the communal practice. *Congruence*: The individual speaker's initial understanding of the term must not diverge so radically from that of others in the community as to undermine that coordinating intention" (Schroeter and Schroeter 2009: 18). On this account, and what makes it distinctive as compared to some other accounts, is that what's required for competence with a moral concept is being causally-historically connected to your linguistic community. This is a weak or at least not a very intellectually demanding view of concept possession. And it's in sharp contrast to some of the more demanding positions that I am not sympathetic to, such as moral functionalism and moral rationalism that require competent speakers to grasp certain canonical propositions (certain moral principles) to qualify as competent (Jackson and Pettit 1995; Peacocke 2004). Whilst almost every metaethical view assumes some very specific theory of concept possession even if proponents don't make it explicit, nothing too substantive in this paper hangs on my commitment to a weak theory of moral concept possession. I merely highlight it at the outset as where I might describe a subject like Abdul as acquiring or improving the skill of concept application, a rationalist or functionalist might deny that the subject has the relevant concept to begin with, and it would be a waste of time for the reader to balk at that point.

Similarly, I also think that Hassan and Abdul can possess some of the same abstract moral principles and rules but apply them very differently. A key motivation for these ideas is that on a standard account of knowledge of concepts and principles on which it is a priori, concepts and principles do not come with a manual on how to apply them.[1] Instead, how to apply them are additional and important skills that have to be learned in experience, where the notion of experience is to be understood very broadly as involving direct real life experiences as well as reading different genres of literature and non-literary writings, viewing artwork, watching films, theatre and so on. However, I am not going to defend these ideas here.

3. Contemporary Accounts of Moral Learning

The second question and answer draw attention to contemporary accounts of moral learning. This section describes some of these development before explaining my answers to the second and third questions more fully in the following section. Cushman, Kumar, Railton (2017) provide a comprehensive summary of developments in accounts of moral learning over the very recent past.[2] We have always known that some kind of learning plays an indisputable role in our moral judgement; anthropological work on the diversity of morality across cultures is one piece of evidence for that idea. Of course, traditionally, theories of moral judge-

1 Morality is not special in this way. As Aristotle points out, the fully skilled soldier will have a theoretical understanding of strategy, but that understanding alone will be no guide in the battlefield. In addition, he must also have a practical appreciation of the risks involved, the potentials for acting, and the risks at stake (*Nicomachean Ethics* 106b23).

2 For detailed discussions of the views covered by Cushman, Kumar and Railton, see the entries in *Cognition 167 (2017)*.

ment have appealed to an innate universal moral grammar or taste buds to explain our moral judgements. But such theories are at best incomplete as accounts of how moral intuitions are learned. Thanks to recent developments in reinforcement learning and Bayesian updating, current plausible accounts of moral learning have utilized the ideas of "model-free" and "model-based" learning to account for patterns in our moral intuitions and intuitive judgements.

Reinforcement learning and Bayesian updating are by their nature expectation-based. They involve the construction of models of the physical and social world, and the models generate expectations that guide perception, thought and action. The model is updated as a result of prediction or expectation error correction. Spatial navigation is the clearest example of reinforcement learning. This is extensively studied in rodents, though there is growing evidence that the same process is found in humans (Shea 2022: 3). Neuroimaging shows that the animal's medial temporal lobe takes various representations – from vision, touch, motion and head direction – to form a cognitive map of the animal's spatial environment. The "perspectival and non-perspectival representations of space and place thus generated guide navigation flexibly by dynamic updating", and what's most fascinating, they can promote offline learning as well (Railton 2017: 174). In particular, when the rodent sleeps, those representations are repeatedly activated and "the trajectories simulated include directions of motion the rat did not experience as well as those it did, and activation focuses especially in regions of the maze less frequently explored – a pattern typical of a learner seeking to extract maximum information from a body of data ..." (Railton 2017: 174). What's more, the rodent constructs in these offline simulations short-cut paths never taken before, thus enriching the map further. The animal's mental map constantly guides it as it navigates the maze.

Evidence suggests that just as with spatial maps, "a cognitive map of social properties was used to infer social relationships between individuals ... where the inferred relationships were not encountered during training, and go beyond chains of associations"; in all these systems "[i]magination is a way of simulating new possibilities within a domain, allowing new connections to be formed or novel categories to be represented" (Shea 2022: 3–4). There is also evidence that prediction or expectation error correction signals in the brain track expectation violations for abstract values, such as uncertainty and untrustworthiness, and not just concrete ones, such as food and mating (Railton 2017: 176).

On Railton's model-based account of moral learning assumed in this paper, it is the result of domain-general learning capacities, rather than a dedicated moral module, and it involves the construction of models of the physical and social world. This modeling, like spatial modeling, generates expectations that guide thought and action and permit learning from discrepancies with expectations, both online and, crucially, offline in simulation or imagination. On Railton's account of "the affective attunement to reasons" as moral learning, the broad affective system is flexible, intelligent (in a sense) and responsive to experience and constructs the aforementioned causal and social models of the world as we navigate it.

Competing dual-process accounts of moral learning include those defended by Greene and Haidt (2002) on which moral intuitions are automatic affective responses evolved because they fall under the kind generally useful to our ancestors. Any arguments we may give in defense of those intuitions (such as in versions of the trolley problem) are post-hoc rationalizations. In contrast, like Railton, Cushman (2012; 2013) has developed an account of moral learning in line with recent work on reinforcement learning, though, importantly, he defends a *model-free* account rather than a model-based one. On model-free learning, prediction error correction builds "cached" expectation values for individual actions or situation-action pairs rather than causal or evaluation models. Model-free learning is demonstrated by everyday examples like unthinkingly taking the same route on a drive one does as a matter of routine even when one knows it is closed because of construction work. This paper assumes the plausibility of a model-based account of moral learning and the next section explains its limitations.

4. Intuitions are Stubborn, Coherence Is Overrated

Having described contemporary accounts of moral learning in the previous section, I will now explain the second and third sets of questions and answers above. Let me begin with Abdul's story.

No Man Is An Island.

Act I

Abdul is a typical male representative of the patriarchal society to which he belongs. He discharges all the duties he associates with his social roles – that of a brother, son, father, officer – with a great sense of responsibility and diligence, in accordance with, and out of respect for, the baggage of values of his socio-economic class, but he is also racist, sexist, misogynistic and so on, in line with the prejudices and biases of the culture in which he was raised.

Through a series of accidents, Abdul lands a job in the administration of a generally egalitarian institution in an urban part of his city. The institution evolved to be egalitarian because of the kind of people it attracts as employees: though raised in the same culture as Abdul, may have travelled widely, had more exposure to world cultures and so on. Thus, to these colleagues of Abdul, whilst his patriarchal ways are familiar, they aren't their ways, and they are not to be encouraged or tolerated in others. Most of his women colleagues confidently interact with men, and many male members of it are feminist in theory and action. There are pockets of conservative corners among his colleagues, and Abdul feels most at home there, but eventually he finds himself out of his depth as the requirements of his job make him venture out into the not-so-conservative circles.

Let me begin at the beginning. Excited about his new job, Abdul is carrying on with his work merrily, ignorant of the moral mistakes common in a patriarchal society. He continues to misconstrue the norms of his new workplace, fitting his colleagues in the usual stereotypes of his society. For example, in his interaction with women at his new workplace who are obviously "modern" (i.e., aligned with the supposedly looser moral values) and, therefore, frivolous and up for grabs according to Abdul's stereotypes, he gets carried away in his crude ways such that his interaction is on the borderline of sexual harassment, though of course he has no realization at all about that fact. The women wince at his sexist jokes, look away when he utters lewd remarks, fidget restlessly when he stares at their bodies, or simply cut him short and walk away, yet initially Abdul sees none of that body language for what it is. Abdul's explanation of this feedback – the women's body language in his presence, the content, tone and timing of their comments to him, and so on – is that the women are playing hard-to-get in the game of flirtation that he imagines to be playing with them; the course of action this conclusion recommends is that he should persevere.

However, as his perseverance is met with an increasingly forceful rejection by the women, slowly but surely Abdul discovers in himself a nagging suspicion that this explanation in terms of playing hard-to-get won't do. Confusion ensues. As time goes on – we are talking not just days or weeks but much longer – the nagging suspicion turns intense, sharp, biting, disposing him to notice how differently they interact with some of their other male colleagues, that those interactions typically last much longer than any he has had with the same women on similar subjects and so on. The inadequacy of his explanation in terms of playing hard-to-get becomes most obvious when one data point shines: that the women walk away from him in not quite the same way as someone in a game of flirtation, but almost as if something in him repulsed them, and he knew he didn't have bad body odours – actually, he smelled quite nice.

As a coping strategy, Abdul starts copying the behaviour and conversational style of those other men he so envies. Gradually, as the imitation becomes better, his revised behaviour gets clear endorsement from the same women who initially avoided him: they no longer walk away from him abruptly. Naturally, this is a prudential gain for Abdul, and it is perceived as such by him. He takes this to be an indication that he has succeeded in learning the social norms of his new workplace, and he continues to switch back to his vulgar and male chauvinistic ways in the familiar contexts. It is as if Abdul is leading a double life.

Act II.

However, as time goes on, Abdul watches the movies recommended by his new colleagues, reads some of the books they recommend and so on. Given this exposure, and as he has more practice at acting as he ought to in his interactions with women, he tends to continue acting that way even in the patriarchal contexts familiar to him, and whilst this results in losses to him, prudentially speaking, it also leads to a new discovery. When he consistently approaches some of his old acquaintances in those patriarchal contexts with his revised behaviour, they show different combinations and degrees of the following surprising reactions: a certain independence in conducting their affairs, better performance at work, school, home, better health, a general display of well-being of which he didn't think they were capable; there are certainly others, though, who incur prudential losses on account of this change and thus react negatively to it. For example, his very own daughter's actions indicate she's becoming more confident; his wife becomes more enterprising, less frequently ill, whereas another relative who is more needy, and manipulative and generally used to taking advantage of patriarchy suffers some losses when Abdul stops indulging her, and so on. None of the described changes always or even mostly work to his advantage – for example, now his word is challenged ever more than before, many of his old friends and relatives, men and women, become hostile towards him, and so on. He notices these reactions and results over a range of cases. Assuming that Abdul has a certain personality type, and that he is not entirely and not always motivated just by selfish interest, his experiences precipitate in a significant shift in his moral evaluations. Yet, it is not at all a complete shift and not at all an easy achievement. Abdul is cognizant of his past errors and, more importantly, of the fact that often, even in the now familiar new contexts, his instincts misfire, as if his old self were always lurking to express itself. The difference is that he is now able to catch those instincts in the act, though sometimes, especially in new and unfamiliar contexts, he has no competing intuitions to guide him. What does he do in these situations? What is the best option for him?

....

The epigraph of Railton's wonderful paper on moral learning, *The Affective Dog and Its Rational Tale,* is the following quote from Shakespeare's *Henry IV:*

Glendower: I can call spirits from the vasty deep.

Hotspur: Why, so can I, or so can any man; But will they come when you do call for them?

I think the answer to Hotspur's question with respect to someone like Abdul is: "not always." The "spirits", the *right* ones, won't come when such a subject calls. The spirits that do come are those stubborn ones entrenched in the "vasty deep" that Abdul's new experience does not penetrate. It's highly unlikely that Abdul will ever fully become like Hassan and Abdul's moral learning is going to look very different from that of Hassan's.

Abdul is an epistemically non-ideal and morally unlucky agent; he is what I will call below a WOTE, that is, a typical member of the wretched of the earth. Abdul is not unique, though he is also not representative of everyone who is non-ideal and unlucky in the way he is. For many WOTEs who are immersed in morally healthy environments as adults may still pick up what's the morally correct evaluation of some situations without worrying whether they have the relevant skill to evaluate unfamiliar, let alone novel, situations accordingly. Such a haphazard change in moral evaluations may qualify as manifesting some degree of moral learning, though not enough and not the kind that a reflective and conscientious agent like Abdul is after, and such a haphazard display of virtue certainly doesn't resemble what Hassan naturally has. The transparency to Abdul of his realization that he is a WOTE[3] is epistemically valuable for his moral learning, or so I shall argue. Moreover, even if it is highly unlikely that Abdul too will fully become like Hassan, he can still achieve moral understanding, though his moral learning will look very different from that of Hassan's.

Why think that Abdul is unlikely to become like Hassan? In the previous section, we saw Railton's affective attunement account of moral learn-

3 Naturally, the content of his knowledge doesn't have to feature "WOTE" or any very sophisticated philosophical concepts.

ing, an account of how the "spirits" are trained in experience: our skill at moral evaluation gets calibrated and tuned by experience. It is a powerful account of that attunement over the course of an individual's life. But what happens if the course changes and the individual switches environments? Railton seems optimistic that "learning in desire" is always responsive to experiential feedback, at least for normal agents. Yet, that idea sounds counterintuitive compared to the common observation that the window for desire to learn, the window for affective attunement, is not open forever and that it in fact closes early on in life after which any change in it will be slow and never deep or complete. This thought is supposed to be preempted in the literature with the help of studies about the change in implicit and explicit attitudes towards homosexuality in the USA over the past few decades (Railton 2014; 2017; Kumar, Kodipady and Young 2023). The key idea is that *implicit* and not just explicit attitudes have changed, supposedly lending support to the idea that our broad affective system can continue to learn new things. The main explanation of the change given by Kumar, Kodipady and Young (2023) is "that many people have discovered that a close friend, family member or admired individual is gay, motivating them to identify the harm and discrimination faced by the individual they know, and catalyzing moral consistency reasoning such that they generalize the interpersonal insight to strangers" (p. 1392). But given that observation, here is an alternative explanation of the change in implicit attitudes: biased individuals' intuitive responses to homosexuality are the *same* as before, but they are now kept in check or outmaneuvered by the competing desire not to let down or betray their gay friends or family members. "What would they think!" one might wonder.

In any case, in the Westgate, Riskind, and Nosek (2015) study, one of the main studies on implicit preferences that the authors refer to, the change in explicit attitudes is *double* the change in implicit attitudes (26% vs 13%), which is not an insignificant fact with respect to the idea that there may be a sensitive window for developing these attitudes and past that window, any changes in implicit attitudes may be at best very slow and probably not deep. It is also noteworthy that the largest proportion of that change in implicit bias in that study is accounted for by the younger members of the population sample "who have lived their entire lives in an atmosphere of more open recognition of sexual orientation" (Railton 2017: 180). In fact, Westgate, Riskind, and Nosek observe that the smallest change in implicit preferences was observed in older adults (among some other categories). So, whilst over all there has been a cultural shift in implicit attitudes towards gay persons in the US, the story at the individual level is very different. There are also some psychological studies that support this point by highlighting the difficulty of completely overcoming prejudices and biases (Devine 1989). Indeed, the phrase "catch them young" might capture a deep psychological fact about us. This is not at all to deny that in more recent work, Devine has been working on designing interventions to reduce implicit biases, or that there are other studies that suggest the possibility of significant success in achieving that reduction. The point, instead, is that complete reversal of biases and prejudices in a single adult is probably not possible. And even if it is, the *transitional* period when one is getting there is, and should be, of interest with respect to moral learning.

If our moral intuitions, like some non-moral intuitions, are shaped early on in life by the kind of experiences we have, and thus not amenable to significant or at least complete revision later in life, moral improvement and understanding will need to recruit cognitive resources other than and in addition to "the spirits" or affect. The thought is that a WOTE would need to watch their moral intuitions and rely on reasoning and reflection much more than a morally lucky agent like Hassan. But wait: doesn't everyone grant the importance of reasoning and reflection in thinking about moral and non-moral matters. Does the current proposal make the banal point that there is just a (great?) difference in the degree to which a WOTE and a Hassan-like agent needs to rely on in moral deliberation? Not really. Instead, the thought is that a WOTE's reliance on reasoning and reflection is not the usual kind discussed in moral philosophy, i.e., reasoning aimed at reaching Rawl's "wide reflective equilibrium" (RE) or its cousin "consistency reasoning" (CR) (Campbell and Kumar 2012). For a non-ideal epistemic and moral agent like a WOTE, RE and CR are likely to have limited use at best. Let me explain why.

According to RE, Abdul should identify and weed out his bad background beliefs because these are not "considered judgements" i.e., "those judgments in which our moral capacities are least likely to be displayed without distortion" (Rawls 1971: 47). Abdul should then check whether the content of a new moral intuition coheres with his considered judgments, retain it if it does, discard it if it does not. Note, however, that considered judgments exist "at all levels of generality, from those about particular situations and institutions up through broad standards and first principles to formal and abstract conditions on moral conceptions" (Rawls 1974: 8). Taking, first, a considered judgement to be a general and abstract moral principle to see whether a particular new moral intuition coheres with it isn't going to help Abdul because he lacks the skill needed to apply abstract principles. Taking a particular moral judgment as the considered judgment will also have limited use given that most particular judgements are, by their nature, context or situation specific. And Abdul is bound to find himself in unfamiliar and novel situations. This also highlights what is problematic about CR. CR is the view that reasoning brings about moral change not primarily through the application of principles, but instead through the identification of conflicts between particular judgements about similar cases making the agent want to revise them to achieve consistency. But, to repeat, particular judgements may be too situation specific. What's more, the crucial step of judging whether two cases are similar might not always be as simple or unproblematic as it is made to sound. So, what is a plausible account of Abdul's moral change and learning?

The notion of a fragmented mind has recently been gaining overdue attention in philosophy (Egan 2008; Borgoni, Kinderman; Onofri 2021). The idea is that traditionally in philosophy, agents are represented as having a single corpus of beliefs which is both consistent and deductively closed and guides all their actions all the time. Instead, the systems of beliefs we in fact have are fragmented. Each belief system is consistent and closed under entailment, but all the systems together need not form a consistent and closed system. It is easiest to describe the key idea in terms of belief systems, but that is just a template for other attitudes, such as knowledge, desire/affect and so on.

Egan (2008) argues that fragmentation is not to be lamented. For a non-ideal agent with a faulty belief forming or a faulty evaluative mechanism, it may be good to be fragmented. "The problem to which fragmentation is a good response is the problem of having the deliverances of some unreliable system fed directly into a system of beliefs that then loses the ability to criticize and discard the faulty input. It's ... bad to be unable to use one's knowledge that one's evaluative mechanisms are unreliable ..." (2008: 61). Hence the importance of Abdul's epistemic self-doubt in his moral learning. Of course, just having inconsistent fragments is not on its own going to be enough to assess a faulty one. What's crucial is whether and how the good fragment gets activated, how the faulty and the good fragments interact and so on.

In *No Man Is An Island*, Abdul is constantly relying on certain heuristics. Once he realizes that his original, WOTE moral sensibility is faulty, whenever he has to morally evaluate a situation involving women, he thinks of a stereotypical WOTE – a type of person he doesn't think he should be – and a Hassan-like, Aristotelian agent – someone he would like to be – and imagines how each of them would evaluate the situation. If his intuitions about the case sit more comfortably with those of the stereotypical WOTE in his imagination, he dismisses them. Even as, with greater experience, Abdul gets affectively better attuned to reasons, he always double checks his intuitive judgments by employing the same heuristics. When his new sensibility draws a blank about a particular case, he imagines the (or the range of) deliverances of his WOTE sensibility, and simply resolves to set those aside because he now knows they are wrong. Thus, Abdul's moral learning relies significantly on his epistemic self-doubt and internal conflict that enable him to access and compare the different fragments of his beliefs, desires and memories.

This is, undoubtedly, an extremely sketchy or preliminary account of the view I have in mind. And it is most certainly a crude characterization of fragmentation. The small amount of literature on fragmentation contains an even smaller discussion of moral attitudes, like implicit biases (see Borgoni, Kirkman, Onofri 2021; Greco 2015; Yalcin (2021). The standard examples concern agents whose behaviour is best explained by their implicit biases, such as that women are less intelligent than men, when their explicit beliefs are egalitarian. Questions arise as to whether we need to invoke fragmentation at all in these contexts to explain the clash in the attitudes of biased agents (see, for example, Toribio 2021; Kinderman and Onofri 2021: 30) even if fragmentation may be plausible for belief states. This is not the place to examine the case against the idea that fragmentation explains the clash of implicit vs explicit attitudes in general. Perhaps it will turn out that the case is plausible. But my assessment so far is that it isn't at least for cases like *No Man Is An Island* which is a special sort of case not considered in the relevant literature on fragmentation. A bad environment early on in life can deeply damage certain cognitive or psychological capacities and the damage may be irreparable just as stunted physical growth due to poor nutrition in childhood cannot be reversed by consuming food fit for royalty later in life. (If Abdul is not so badly damaged, he just might become like Hassan one day, there is still the transitional period to go through.) Once immersed in healthier moral environments as adults, internal conflict between intuitions generated by competing sensibilities is not a matter of implicit vs explicit attitudes. In the best case scenario, the same situational features trigger or give the agent access to all (both) the competing sensibilities – what I have called *fragments* – and it is by comparing their workings that the agent settles on which on intuition is more reliable. The comparison (and the epistemic self-doubt that makes it possible) plays a key role in Abdul's moral learning, and that's why remaining fragmented is good for him. It may sound crazy to traditional moral epistemologists that fragmentation of mind sometimes plays such an instrumental role in moral learning and thus needs to be guarded, but "though this be madness, yet there is method in it"[4]. This picture does not at all rely on the Cartesian assumption of the luminosity of the mental, but just that a WOTE needs to rely on introspection in situations where a moral judgement is to be made, that they simply need to put themselves on the proverbial Freudian couch when matters of value make a claim on their attention.

Conclusion

It is a standard assumption in philosophy that the more consistent and coherent an agent is, the better for them. However, this idea relies on the assumption that we are all ideal rational agents lucky enough to be in a healthy moral environment. The most plausible account of moral learning today identifies mechanisms that do not differentiate among different kinds of environment one is in: we can become affectively attuned to whatever are taken to be reasons in the environment. But "learning" is a success term. How does one gain moral knowledge or understanding if one is, for example, a WOTE raised in a defective moral environment? Retraining the broad affective system is very hard beyond a certain age. This paper has argued that a conscientious moral agent like Abdul, who is also fortunate enough to realize that their moral environment is defective, can make good use of their epistemic self-doubt and internal tension generated by conflicting intuitions to marshal cognitive resources other than affect, and capitalize on the fragmentation of his mind, to gain moral understanding.

References

Adolphs, R. & Anderson, D.J. 2018. *The Neurobiology of Emotion: A New Synthesis.* Princeton University Press.

Aristotle. 2000. *Nicomachean Ethics*, ed. and trans. Roger Crisp. Cambridge: Cambridge University Press.

Campbell, R. and Kumar, V. 2012. "Moral Reasoning on the Ground." *Ethics 122* (2), 273 – 312.

Cristina Borgoni, Dirk Kinderman and & Andrea Onofri (eds.), 2021. *The Fragmented Mind*. Oxford: Oxford University Press.

Cushman, F. A. 2013. "Action, Outcome and Value in a Dual-Process Framework for Morality." *Personality and Social Psychology Review, 17,* 273–292.

Damasio, A. 1994. *Descartes' Error.* New York: Avon Books.

4 Shakespeare, *Hamlet*

———, 2003. *Looking for Spinoza.* Harcourt Books

———, 2021. *Feeling and Knowing*. New York: Pantheon Books.

Devine, P. 1989. "Stereotypes and Prejudice: Their Automatic and Controlled Components," *Journal of Personality and Social Psychology* 56: 5–18.

Egan, A. 2008. "Seeing and Believing: Perception, Belief Formation and the Divided Mind." *Philosophical Studies* 140 (1):47 – 63.

Greco, D. 2015. "Iteration and Fragmentation." *Philosophy and Phenomenological Research* 91/3: 656–673.

Greene, J. and Haidt, J. 2002. "How and Where does Moral Judgement Work?" *Trends in Cognitive Science, 6,* 517 – 523.

Hume, D. 1739–40/2000. *A Treatise of Human Nature*. David Fate Norton and Mary J. Norton(eds). Oxford: Clarendon Press.

Jackson, F. and Pettit, P. 1995. "Moral Functionalism and Moral Motivation." *Philosophical Quarterly* 45, 20–40.

Jackson, F. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. New York: Oxford University Press.

Jones, K. 2006. "Quick and Smart? Modularity and the Pro-Emotion Consensus." *Canadian Journal of Philosophy* (36): 3–27.

Kumar, V., Kodipady, A. and Young, L. 2023. "A Psychological Account of the Unique Decline in Anti-Gay Attitudes." *Philosophical Psychology 38*(4), 1391–1425.

May, J. 2021. "Moral Rationalism on the Brain" *Mind and Language.* 23

Hensch, T.K. 2004. "Critical Period Regulation," *Annu Rev Neurosci*. 27:549–579.

Knight, Carl, "Reflective Equilibrium", *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.),

URL = <https://plato.stanford.edu/archives/win2023/entries/reflective-equilibrium/>.

Mladinow, L. 2022. *Emotions: How Feelings Shape Our Thinking.* New York: Pantheon Books.

Mobbs, D. et al., 2019. "Viewpoints: Approaches to Defining and Investigating Fear," *Nat Neurosci* Aug 22(8): 1205– 1216

Peacocke, C. 2004. *The Realm of Reason*. Oxford: Oxford University Press

Panksepp J., 1998. *Affective Neuroscience*. New York: Oxford University Press.

Railton, P. 1986. "Moral Realism." *Philosophical Review.* 95: 163– 207.

———, 2010. "A Unified Account of Rationality in Belief and Action". Unpublished manuscript online.

———, 2014. "The Affective Dog and Its Rational Tale: Intuition and Attunement." *Ethics* 124(4):813–859.

———, 2017. "Moral Learning: Conceptual Foundations and Normative Relevance." *Cognition* 167 (C):172–190

———, 2018. "Learning and Doing." *John Locke Lectures, Oxford.*

Rawls, John, 1971, *A Theory of Justice*, Cambridge, MA: Belknap Press of Harvard University Press.

———, 1974, "The Independence of Moral Theory", *Proceedings and Addresses of the American Philosophical Association*, 48: 5–22. doi:10.2307/3129858

Shea, N. 2022. "Concepts as Plug and Play Devices." *Phil. Trans. R. Soc. B*37820210353.

Shroeter, L. and Shroeter, F. 2009 "A Third Way in Metaethics." *Noûs* 43 (1):1–30.

Toribio, J. 2021. "Implicit Bias and the Fragmented Mind." In Cristina Borgoni, Dirk Kinderman and & Andrea Onofri (eds.), 2021. *The Fragmented Mind*. Oxford: Oxford University Press.

Westgate, E. C. Riskind, R. G. and Nosek, B. A. 2015. "Implicit Preferences for Straight People Over Lesbian Women and Gay Men Weakened from 2006 to 2013. *Collabra 1,* 1 –10.

Woodard, K. and Pollak, S. D., 2020. "Is There Evidence for Sensitive Periods in Emotional Development?" *Curr Opin Behav Sci* Dec, 36: 1–6.

Yalcin, S. 2021. "Fragmented but Rational." In Cristina Borgoni, Dirk Kinderman and & Andrea Onofri (eds.), *The Fragmented Mind*. Oxford: Oxford University Press.

https://doi.org/10.18485/baf.2025.2.ch11

Marcin Trepczyński

VIRTUES AS DISPOSITIONS: DIFFERENT APPROACHES IN MEDIEVAL ANALYTIC THOUGHT

Abstract: Is virtue a disposition? Can we understand the Greek "hexis"—used in the Aristotelian definition of virtue—as a kind of disposition? Were virtues considered dispositions in medieval philosophy and theology? I am going to show that definitions of virtues represent different approaches with respect to the first question and that medieval analytic thinkers who developed interesting theories concerning virtues and the concept of *habitus*, were not unanimous in this matter. In my paper, I refer to works of such figures as Peter Abelard, John of Salisbury, Alain of Lille and Simon of Tournai, as well as to fragments from recently edited *Quaestiones Theologiae* by Stephen Langton. Finally, I am going to present the theory of Thomas Aquinas. The analyses provided by these authors reveal that disposition was conceptualized differently by medieval analytic authors, and that this concept did not always fit with the definition of virtue. They also give us valuable insights concerning the notion of disposition.

Key words: virtue, *habitus*, disposition, Aristotle, Peter Abelard, John of Salisbury, Simon of Tournai, Stephen Langton, Thomas Aquinas

1. Introduction

Can we say that virtues, e.g. courage, temperance, friendliness or justice, are dispositions? This question seems important if we try to define the range of our potential dispositions. Virtues are natural candidates to be part of them, due to the fact that in some translations of Aristotle's *Nicomachean Ethics*, they are explicitly defined as dispositions. However, one can doubt: can we understand the Greek "hexis" used in the Aristotelian definition of virtue as "disposition"? There is also another important question: is the Aristotelian approach to the concepts of virtues and dispositions the only correct way to understand them?

An interesting set of material which is helpful in considering these problems are several utterances of medieval thinkers, who for their precision and analytic approach, represent what can be called medieval analytic thought. Their analyses and insights can inspire us to consider different possible relationships between such concepts as: virtue, *habitus* and disposition. In their examples, we can see an evolution of thinking about those relationships and examine the roots of a shift in understanding the term "disposition."

In this paper, I will first refer to a seminal article on the concept of habitus by Cary J. Nederman (1989), who examined the relationships between habitus and virtue in the works of prominent medieval thinkers, and to a polemic text by Marcia L. Colish (1993). These two papers provide us with deep knowledge on how medievals understood and used such terms. However, although these and other authors in different works dealt with medieval definitions of virtue which referred to the concept of disposition (e.g. Cunningham 1969), among them many works on Aquinas's theory of virtue (e.g., Torchia 2002; Goris 2017; Whitmore 2018; Mattison III 2019; Swanstrom 2020), or asked about virtues as dispositions in a modern context (Sockett 2009; Azzano and Raimondi 2023), to my knowledge, they have not directly posed and answered the question of whether medieval authors, including early scholastics, understood virtues as dispositions. I hope this article will be a valuable contribution to the debate.

Below, I plan to refer to an interesting fragment from Peter Abelard's *Ethics*, and compare it to the approach of his disciple John of Salisbury, adding a critical note to one of Nederman's interpretations. Next, I will present the approach of Simon of Tournai, as well as interesting fragments from recently edited and published *Quaestiones Theologiae* by Stephen Langton, and conclude with the theory provided by Thomas Aquinas. This material will be very helpful in answering the questions posed in the first paragraph of this introduction. Its presentation will be preceded by a comparison of definitions of virtue, including different translations of the definition provided by Aristotle.

Lastly, let me add a short linguistic note: I will use English words "virtue" and "disposition," but at the same time I will write "*habitus*" and not "habit" to avoid misunderstanding caused by the modern use of this term. This decision has been inspired by a caveat, which will be quoted in section 6, made by Nicholas Austin (2017, p. 23–24).

2. Definitions

Let us start with selected modern and medieval definitions of virtue expressed in English. A relatively representative example of the modern approach to this term has been presented by Rosalind Hursthouse and

Glen Pettigove (2023) in the entry "Virtue Ethics" in the Stanford Encyclopedia of Philosophy. The authors define virtue as "an excellent trait of character" and "a disposition, well entrenched in its possessor—something that, as we say, goes all the way down, unlike a habit such as being a tea-drinker—to notice, expect, value, feel, desire, choose, act, and react in certain characteristic ways." Hence, they explicitly state that virtues are dispositions. They also contrast them with "habits," understood, as we will see below, in an absolutely different way than medieval thinkers did. Contrary to a virtue, a habit would be something which is not "well entrenched in its possessor." An illustration they give is being a tea-drinker, so we can deduce that, in their view, being a tea-drinker is something that is not deeply rooted in a person and that can be easily changed.

So how did medievals define virtue, and how did they understand such terms as "*habitus*" and "disposition"? As Jeroen J.W. Laemers (2005, p. 119) points out, a current definition of virtue in the twelfth century was: "virtue is the condition or disposition of the well-ordered mind" (*virtus est habitus mentis bene constitutae*). He indicates that this formula was present in the writings of Boethius and that it was him who introduced "this definition into the ethical discourse of the Latin West." He also reports that both theologians (he mentions Anselm of Laon and Peter Abelard) and monastic writers, as well as civil lawyers adopted this formula. The examples analysed below will confirm his thesis. We should also note that Laemers decided to translate the word "*habitus*" as "condition or disposition." This interpretation decides that virtues are or can be understood as dispositions. Such a translation decision is not illegitimate. Yet, as we will see below, at least some twelfth-century thinkers would not be able to accept it, if we assume that "disposition" means the same as what they called *dispositio*.

Notably, Nederman translates this word differently. As he argues at the beginning of his article, "among the range of moral concepts that the Middle Ages derived from Aristotle, few exercised greater influence than the doctrine of *habitus* (a term ordinarily translated as 'habit,' but more properly meaning 'state' or 'condition')" (Nederman 1989, p. 87). Hence, "condition" appears again, but instead of "disposition" Nederman proposes the word "state." We should note that in the next sentence, he supports his claim by pointing out the thirteenth-century thinkers, such as Thomas Aquinas, Godfrey of Fontaines, Duns Scotus, and William of Ockham, who "placed *habitus* (derived from the Greek term ἕξις) near the heart of their studies of ethics." He also adds that this "thirteenth-century interest in the concept of *habitus*" is likely to be explained "on the basis of the appearance of Robert Grosseteste's full translation of Aristotle's *Nicomachean*

Ethics." It clearly shows that Nederman combines the career of this word in the ethical discourse with the entrance onto the scene of Aristotle's ethics. He adopts the Aristotelian understanding of *habitus* and presents it as an equivalent of Greek ἕξις. However, it will not apply to the use of the word *habitus* by the twelfth-century thinkers, as Colish will later point out (1993, p. 80).

Let us finally take a look at Aristotle's definition of virtue and its English translations. As Stagirite writes in his *Nicomachean Ethics*:

> ἔστιν ἄρα ἡ ἀρετὴ ἕξις προαιρετική, ἐν μεσότητι οὖσα τῇ πρὸς ἡμᾶς, ὡρισμένῃ λόγῳ καὶ ᾧ ἂν ὁ φρόνιμος ὁρίσειεν. (Aristotle, *Ethica*, II.6, 1106b36–1107a1)

According to the translation of David Ross from 1925:

> Virtue, then, is a state of character concerned with choice, lying in a mean, i.e. the mean relative to us, this being determined by reason, and by that reason by which the man of practical wisdom would determine it. (Aristotle 1925, p. 31)

Greek ἕξις has been translated as "a state of character." In the translation of Jonathan Barnes from 1984, it is shortened to "a state." The first part of this definition is translated as follows: "Excellence, then, is a state concerned with choice... (...)" (Aristotle 1984, electronic markup 1991, p. 25). Both translations emphasise that, before all, virtue is something stable that exists in a human and that is a fact concerning this person. This is an approach which resigns from the word "disposition."

By contrast, the translation by Harris Rackham, first published in 1926, reconciles the above-mentioned stability with the concept of disposition. According to this translation: "Virtue then is a settled disposition of the mind determining the choice of actions and emotions... (...)" (ed. Aristotle 1956, p. 95). So ἕξις is a disposition, however, a special kind of disposition, such that it is settled in humans, which corresponds to the idea of stability. And consequently, virtue is a kind of disposition as well.

Also Richard Kraut, in his entry "Aristotle's Ethics" in the Stanford Encyclopedia of Philosophy, includes "disposition" as an element of the definition of virtue and of the meaning of ἕξις. As he states, "Aristotle describes ethical virtue as a 'hexis' ('state' 'condition' 'disposition')—a tendency or disposition, induced by our habits, to have appropriate feelings" (Kraut 2022). Interestingly, he distinguishes disposition and what he calls "habits" as two elements, such that the former is induced by the latter. At the same time, disposition is one possible meaning of ἕξις. Thus, the "habits" that he mentions should not be identified with *habitus* which is a Latin word for ἕξις. Anyhow, what is most important at this stage is that Kraut represents the approach according to which, in Aristotle's view, virtue can be understood as disposition, and presents it in one of the most influential philosophical encyclopedias.

3. Peter Abelard and John of Salisbury

Peter Abelard precisely set up for the next generations the relationship between the three concepts: virtue, *habitus* and disposition. In his theory, he included two important elements of this relationship which are: the process of gaining virtues and the degree of their stability.

First, as Nederman notices, according to Abelard, "virtue is something which is necessarily the product of human effort" (Nederman 1989, p. 95). This scholar points out that for Abelard, shaping such a moral character requires "struggle" and "active process." "Purely natural or instinctual inclination" is not sufficient to speak about virtue. Even more, in Abelard's view, virtue can be credited to someone who develops his character "contrary to his nature" (Nederman 1989, p. 96).

Second, Abelard utilizes the above-mentioned formula transmitted to the Latin West by Boethius and clearly distinguishes between *habitus* and dispositions. Let us quote an extensive fragment from his *Ethics*:

> The philosophers have agreed that virtue is by no means said to be in us unless it is a *habitus* of the best mind or a *habitus* of the well-ordered mind. Now, what they call *habitus* or *dispositio*, Aristotle diligently distinguished in the first species of quality, that is, teaching that those qualities which do not inhere naturally in us, but arrive by our application, are designated *habitus* or *disposiciones*. They are called *habitus* if they are difficult to change – such are, he says, knowledge and the virtues; by contrast, they are called *disposiciones* if they are easily changed. (Abelard, *Ethica*, 128, transl. Nederman 1989)

Hence, we see that Abelard knows Aristotle's idea of *habitus* and dispositions. He shows that both of them belong to the category of quality (according to the Aristotelian framework of ten categories) and are not natural, but require application, or in other words: a sort of practical training. At the same time, there is a huge difference between them. Disposition can be easily changed, whereas *habitus* is quite the opposite. Finally, Abelard expresses a common opinion of philosophers that we have virtue if it is a *habitus*. Hence, *a contrario*, virtue is not a disposition.

We should note here that Aristotle indeed presents such a distinction in his *Categories*, where he writes:

> One sort of quality let us call 'habit' or 'disposition'. Habit differs from disposition in being more lasting and more firmly established. The various kinds of knowledge and of virtue are habits... (...) By a disposition, on the other hand, we mean a condition that is easily changed and quickly gives place to its opposite. (Aristotle, *Categories*, 8, transl. Edghill)

However, a few sentences later the Stagirite adds:

> Habits are at the same time dispositions, but dispositions are not necessarily habits. For those who have some specific habit may be said also, in virtue of that habit, to be thus or thus disposed; but those who are disposed in some specific way have not in all cases the corresponding habit. (ibid.)

This means that *habitus* is also a disposition. However, the two are distinct, because not every disposition is a *habitus*. Hence, only a disposition which is not a *habitus* can be easily changed, not every disposition. However, it seems that this fact was overlooked or neglected by medieval thinkers. Importantly, they also did not know Book V of *Metaphysics*, which would come onto the scene in the 13th century and which closely connects these two concepts. Before that time, a strong opposition between them had been established. And consequently, virtue was not considered a disposition, but only *habitus*.

Abelard's disciple, John of Salisbury, follows the approach of his master and adds some new insights. Nederman (1989, p. 100) indicates that "John explicitly refers in this regard to the Aristotelian distinction between διάθεσις (*dispositio*) and ἕξις (*habitus*)." He points out John's utterance according to which theoretical principles "must be consolidated by practice and exercise, except perhaps where a disposition has already been transformed into a *habitus*."[1] According to this statement, disposition seems to be a step towards *habitus*. It can be understood as a stage which precedes *habitus*. Hence, disposition and *habitus* are not just opposed to each other. They are distinct, but at the same time they are connected as two stages of one process.

Let us add that Nederman claims that John presents the end of this process as the formation of a "second nature." He describes John's proposed "second nature" (*altera natura*) as "an irresistible but acquired inclination to behave in accordance with definite precepts." As he explains, it is "present only when a *habitus* has been engendered, that is, only when a

1 John of Salisbury, *Metalogicon* (*Patrologia Latina* 932c): "(...) si non usu et exercitio assiduo roboretur: nisi forte in habitum transient dispositio."

quality becomes so thoroughly person that it inheres in him." (Nederman 1989, p. 100–101).

However, in my opinion, Nederman's presentation is a bit exaggerated and lacks sufficient textual evidence. This author does not provide any quotation when mentioning the connection between *habitus* and second nature (Nederman 1989, p. 102) and just notes in a footnote that "the notion of an *altera natura* is discussed at *Policraticus* 489b and in *Letters* II 144." Unfortunately, in the *Polycraticus* we can only find one sentence in which John says that according to some people (so it is an opinion he revokes) use (*usus*) is difficult to unlearn, and custom (*consuetudo*) assists another (or: second) nature (*altera natura*).[2] Next, the only place where John mentions the two types of nature (*primitiva natura* and *altera natura*) is in his letter to Bartholomew, a bishop of Exeter. However, he speaks there about primitive nature which has been joint in Christ with divine nature. John explains that he calls it primitive to distinguish it from another nature which corresponds to our degenerate state according to which we are children of anger. He mentions "another nature" as part of the reference to Cicero: he says that according to the Orator, use (*usus*) is another nature (*altera natura*), from which it is very difficult to be torn away.[3]

Hence, John of Salisbury does not introduce the idea of *altera natura* and does not combine it with the concept of *habitus*. He just borrows this idea from Cicero. In the latter case, he absolutely does not relate it to the process of our formation. Quite the opposite: he presents it in a negative way, in the context of our degeneration. However, on these grounds, he indeed proposes the idea of *primitiva natura*, as our pure human nature, not influenced by our customs and abuses, as distinct from the domain of *usus*, which constitutes another nature.

Yet, even if Nederman's presentation goes too far, it seems that John could accept using this distinction in the way Nederman proposed. So, it is likely that the philosopher from Salisbury would agree that virtues are part of such another/second nature. However, although not without serious grounds, it is only a supposition. What we can say for sure, is that

2 John of Salisbury, *Polycraticus* (*PL* 489b-c): "Usus enim, ut ait quidam, aegre dediscitur, et consuetudo alteri naturae assistit."

3 John of Salisbury, *Epistola CLXXXIV. Ad Bartholomaeum Exoniensem Episcopum* (*PL* 189a-b): "Sunt enim corpus unum, (...), dum admirabili commercio illa, quae carnis sunt ex natura primitiva Domino impertit, ut ab eo plenitudinem divinae naturae recipiat (...). Primitivam naturam dixerim ne abusionis inveteratus mos natura reputetur, juxta quem omnes sumus natura filii irae, non quod in ea conditi sumus, sed quia in eam degeneravimus. Nam ut ait orator, usus altera natura est, a quo difficillimum est avelli."

John, who followed Abelard, like other generations of philosophers, including Albert the Great (see Cunningham 1969, 114), improved his theory. He showed that although *habitus* is different from disposition, at the same time *habitus*, and in consequence, virtue, can arise from disposition as a result of consolidation, which makes it resistant to change.

4. Alan of Lille and Simon of Tournai

Nederman emphasises a statement of Alan of Lille and Simon of Tournai, according to which "the genus of virtue is the category of quality."[4] Although Abelard already stated the same, Nederman argues that "this claim is significant in helping to determine when virtue can properly be ascribed to an agent" (Nederman 1989, p. 104).

In my view, Simon in fact repeats what we find in Abelard's and John's utterances, although he puts it differently. He says that such a quality as disposition, which is easily changeable, is not called virtue, but it happens when it is changed (*versa*) in a tenacious *habitus*. However, there is at least one aspect which is worth mentioning. Simon stresses more on the time-perspective. He concludes that virtue does not seem to be something momentaneous, but durable.[5] Hence, disposition, contrary to virtue which is a *habitus*, can be understood as a quality which is transient and may easily disappear.

Importantly, it does not exhaust Simon's contribution to the issue we are discussing here. Simon provides us with another important insight concerning the concepts of virtue and *habitus*, probably overlooked by Nederman. I mean his considerations on virtue in the perspective of the distinction between two aspects expressed in Latin as: *in habitu* and *in usu*.

These considerations regard a theological thesis concerning the virtues according to which "he who has one virtue has all of them," which is connected to the problem of interconnection of virtues (see more: Bieniak 2014). Simon emphasises that we should understand it as "he who has one virtue as *habitus* has all of them." Consequently, "also he who is incapable of exercising one, is incapable of exercising any of them." However, as he points out, "by virtue we do not call only something which we have or a *habitus*, but also use (*usus*)." In the latter sense, we cannot admit that he who has one of virtues, has all of them, Simon claims. Similarly, he shows

4 "Est enim qualitas genere predicamenti" (after: Lottin, *Psychologie et morale* III ii 107; cf. Alan of Lille, *Tractatus* 47).

5 Lottin, *Psychologie et morale* III ii 106: "Que qualitas quamdiu est dispositio facile mobilis non dicitur virtus, sed tunc demum cum versa est in tenacem habitum. Unde virtus non videtur momentanea, sed diuturna."

that vice can be considered as *habitus* (as something contrary to virtue as *habitus*) or as abuse (contrary to virtue as *usus*).[6]

Colish indicates the significance of the distinction between virtue *in habitu* and *in usu*, to criticize Nederman for presenting Alan and Simon "as full-blooded Aristotelians" (Colish 1993, p. 80). As she argues, "this formula suggests a very non-Aristotelian understanding of the term, whereby *habitus*, as inner intention, is contrasted with *usus*, as its outer expression." Colish also refers to Artur Michael Landgraf, who discovered an unpublished mid-twelfth-century manuscript which ascribes this distinction to Peter Lombard. And as she indicates:

> the Lombard can certainly be classified as one of those thinkers who, when they use the term *habitus*, do not feel a need to view it according to the precepts of any one philosophical school. In his own definition of virtue as a *habitus mentis* disposed toward the good, for instance, he does not indicate whether he means this term as a fixed ethical intentionality along the lines of the Stoics, a mental disposition only, or a more Aristotelianizing ethical intentionality developed and reinforced by exercise. (ibid.)

This constatation warns us not to use Aristotelian interpretation as the only key to understand the concept of *habitus* in twelfth-century thought. As we saw, it refers also to Simon, who employed the distinction: *in habitu* vs. *in usu*. It is then natural to see how this distinction and the pair disposition-*habitus* work in texts of another prominent Parisian master of those times, whose *Theological Questions* has recently been edited and published, namely Stephen Langton.

5. Stephen Langton

Stephen Langton encounters the distinction *in habitu* vs. *in usu* when he discusses a question of chastity in the example of Abraham. In his question 89 *On virginity*, which is part of his *Theological Questions* (see Stephen Langton 2022), he refers to the following statement of St. Augustine, which he finds in the *Sentences* of Peter Lombard: "the chastity of the unmarried is better than the chastity of marriage; whereof

6 Simon of Tournai, *Disputationes*, d. 33, q. 1, p. 97: "Sicut autem qui habet unam virtutem habitu, habet omnes ut dictum est, ita quoque qui inhabilis est ad exequendum unam, et ad quamlibet exequendam inhabilis est. Unde si vitium intelligitur habitus vel habilitatis privatio, qui unum vitium habet, et omnia. Sed virtus non solum haberi dicitur natura vel habitu, sed etiam usu, ut dictum est. Quo modo non qui unam habet, et omnes. Usui vero virtutis contrarius est abusus, qui dicitur vitium secundum quod soliti sumus dicere."

Abraham had one in use, both in habit."[7] It is clear that Abraham could perform chastity of marriage, which is represented by chastity *in usu*. But how does Langton explain that Abraham had both forms of chastity *in habitu*?

In one of three versions of question 89 *On virginity*, marked as q. 89g, Langton points out, in line with Augustine's solution, that Abraham was capable (*habilis*) of both forms. He argues that if God willed, Abraham would also be happy to perform the chastity of the unmarried. Finally, Langton shows that *habitus* can be understood here in different ways. Although Abraham did not have virginity in such a way that he could be called a virgin, he had it in a way that if his wife died, he would have widower's continence, Langton argues.[8]

In this example, we see that Langton applies the discussed distinction and that *habitus* can be understood in multiple ways. In general, Langton reads *habitus* as an ability or capacity of a person. So, such a capacity may be understood on different levels: a capacity which can be or is being performed, or a capacity which a person has, which could be performed in different circumstances. As he shows, the virtue of continence is such a *habitus* that can be considered in different ways. It also seems that, in line with Colish's observation, this approach to *habitus* is not necessarily based on the Aristotelian concept.

Another example that illustrates utilizing the concept of *habitus*, connected to the distinction *in usu* vs. *in habitu*, can be found on q. 90. Similarly to Simon of Tournai, Langton applies this distinction to the discussion of the controversial statement according to which he who has one of the virtues has all of them. Langton finally accepts this thesis and argues that a person who has virtues is like a king who fights with bad spirits and carnal temptations using his army of virtues. Although he uses particular virtues for particular enemies, each virtue is ready to help and defend. He

---

7 Augustine, *On the Good of Marriage*; cf. Augustine, *De bono coniugali*, c. 22, 27, ed. Zycha, CSEL 41, 222 (*PL* 40, 392): "melior est castitas caelibum quam castitas nuptiarum: quarum Abraham unam habebat in usu, ambas in habitu." Peter Lombard, *Sententiae*, d. 33, c. 2, 1 (Roma, Grottaferrata 1971, II, p. 460).

8 Stephen Langton, *Quaestiones Theologiae* (hereinafter: *QT*), q. 89g, c. 8, p. 348: "Set ibi procedit quod "Abraham habuit unam in usu et ambas in habitu". Et notandum quod continentia ibi duobus modis accipitur, scilicet large et stricte. Utramque dicitur habuisse quia habilis fuit ad utramque, quia si deus uellet, tam libenter haberet hanc ut illam. Et diuersimode accipitur hic 'habitu', quia non habuit uirginitatem ita in habitu ut ab ea posset denominari 'uirgo'. Vel habuit in habitu quia, si <uxor eius> moreretur, haberet continentiam uidualem." Cf. ibid., q. 89b, c. 9.2, p. 342: ""In habitu", idest habilis fuit ad continendum et paratus continere quantum lex matrimonii permisit."

concludes that this means precisely "to have virtues *in habitu*."[9] Although they are not being used at once or currently, they are ready to be used whenever they are needed.

However, Langton also provides us with insights concerning the concept of *habitus* as compared to the notion of disposition. He refers to them in all three versions of q. 89, marked as q. 89b, q. 89c and q. 89g. We find there a distinction similar to the one we know from previously mentioned authors. In q. 89b, Langton states "that to be virginity is for virginity something accidental, similarly, for a disposition to be a disposition, because it will be a disposition when it will not be a disposition, but a *habitus*." As he affirms, after a sexual encounter, if there was no mortal sin, the virtue remains, but this virtue which was "virginity or virginal continence, is now a conjugal chastity."[10]

Does he say anything new here? In my opinion: yes. Apart from distinguishing disposition from *habitus*, he emphasises a strong connection between them. He not only shows that a disposition can be transformed into a *habitus*, like John of Salisbury indicates; he also seems to stress that it is still the same thing, although in some way different. To point out this identity, he uses a contradictory formula, according to which a disposition will be a disposition and at the same time will not be a disposition. It seems that he tries to say that *habitus* is still a disposition, but the one which became a *habitus*. It is the same thing, but modified, namely strengthened.

In q. 89g, he does not use the contradictory formula, but speaks about one and the same property, which "is now a disposition, and will be a *habitus*." Hence, although he does not give space to say that *habitus* is disposition, he emphasises that both of them are one property, but in different states, similarly to a virtue which once is virginity, and next is no longer virginity, although it is still this virtue.[11]

---

9 Stephen Langton, *QT*, q. 90, c. 6, p. 357: "ita et homo habens uirtutes est quasi rex qui congreditur contra spiritus malignos et carnis temptationes cum exercitu uirtutum, quarum quedam percutiunt, ut continentia spiritum luxurie si temptatur de luxuria, alie autem uirtutes parate sunt ut pugnent si opus fuerit, ut si temptetur de superbia, pugnet humilitas etc. (...) Oportet enim ut una uirtus suffragetur alii, et hoc est habere uirtutes in habitu, idest habere eas paratas ad defendendum si opus fuerit."

10 Stephen Langton, *QT*, q. 89b, c. 1, p. 339: "Accidentale est uirginitati esse uirginitatem, sicut dispositioni esse dispositionem, erit enim dispositio quando non erit dispositio set habitus. Quod patet in uirgine contrahente: ante contractum fuit in ea uirginitas, post cognitionem non, et tamen uirtutem quam habuit non amisit, quia uirtus nonnisi per peccatum mortale amittitur. Illa ergo uirtus que fuit uirginitas uel continentia uirginalis modo est castitas coniugalis."

11 Stephen Langton, *QT*, q. 89g, c. 1, p. 344: "Accidentale est uirtuti esse uirginitatem, quia erit uirtus cum iam non erit uirginitas. Simile potes habere in dispositione et habitu, quia hec proprietas que modo est dispositio iam erit habitus."

Let us also mention Langton's utterance in q. 89c, where he also does not use the contradictory formula, but adds an element known from Abelard's utterance. Langton states "that to be virginity is for virginity something accidental, because it can be also conjugal or widow's continence." He asserts that, similarly, "it is accidental for a disposition, which is easily changeable, to be a disposition, because it can be a *habitus*, which is hard to change."[12] Hence, this time he follows the distinction adopted by philosophers mentioned by Abelard, which is based on the criterion of changeability.

In all three versions, Langton presents the analogy between a virtue of virginity and disposition to argue that virginity is not something that can grow in a way that someone will become a greater virgin. We cannot say that when a disposition grows, one who has it will be more disposed, because, as Langton says in q. 89c: "I assume that the disposition would become a *habitus.*"[13] At the same time, he emphasises that disposition is not something that can grow in such a way that it will extend itself or will get more intensive, e.g., that it will give more power. Similarly, as he argues, a young man who grows will not get younger. When a disposition grows, it gets firmer and finally becomes a *habitus*.

It is also worth mentioning that Langton applies the notion of *habitus* to the discussion on love. When analysing the problem of intensity of love in the example of a sentence "this person loves more God than a neighbour," in one of arguments he introduces three meanings of loving. As he indicates in q. 76.2, "the word 'loving' refers to some *habitus* between love (*caritas*) and the movement,"[14] where by movement he understands action or act of loving. Or as he puts it in other words: "the word 'loving' can be understood in three ways: sometimes it refers to love, sometimes to a *habitus* or a capacity (*facultas*), sometimes to a movement of love." Thus, when we say that someone sleeping loves God more than a neighbour, "loving" refers to such a capacity. Moreover, we could say that there are two faculties of loving: loving God and loving a neighbour, with different intensities, though they come from one love.[15]

---

12 Stephen Langton, *QT*, q. 89c, c. 1, p. 334: "Accidentale est uirginitati esse uirginitatem, quia potest esse coniugalis continentia uel uidualis, sicut accidentale est dispositioni, que est facile mobilis, esse dispositionem, quia potest esse habitus, qui est difficile mobilis."

13 Stephen Langton, *QT*, q. 89c, c. 1, p. 334: "Uel sic: dispositio est in isto, ergo ea crescente erit magis dispositus quam sit, quod falsum est, quia pono quod dispositio fiat habitus."

14 Stephen Langton, *QT*, q. 76.2, c. 1.1, p. 159: "uerbum diligendi copulat quendam habitum medium inter caritatem et motum."

15 Stephen Langton, *QT*, q. 76.2, c. 1.1, p. 159–160: "Et uerbum diligendi tripliciter dicitur: quandoque copulat caritatem, quandoque habitum siue facultatem, quandoque motum caritatis. Et ita, cum dicitur 'iste dormiens magis diligit deum

However, although Langton decides that such a solution to the problem could be accepted, he ultimately adds nuance to this theory. He says that loving refers to love (*caritas*), which is one, whereas the intensity of love refers to its effects, which are its movements, however: not the movements which someone has (because a sleeping person does not have such movements), but the ones which are *in habitu*. This time Langton gives an example of a sentence "God loves more this person than that" to show that "loving" does not refer here to the love in God (as in the case of God, his love is his essence, so it must be one), but to the intensity of effects of love in different persons. He means effects which these persons have or will have, and probably for this reason he calls it a movement of love *in habitu*.[16] This case again illustrates the different uses of the word *habitus* in Langton's theological analyses. Sometimes *habitus* is almost the same as a capacity (*facultas*), and in other situations it is something that someone can have as an effect of someone else's action. Obviously, virtue is not a *habitus* in the latter sense.

Lastly, let us mention that Langton applies the concept of *habitus* to the gifts of the Holy Spirit. He considers the question of whether such gifts are different from virtues and whether they are prior to them, simultaneous, or posterior. He decides that "the gifts are posterior to virtues, because they are effects or quasi-effects of virtues." Next, he adds that by "effect" he does not mean the use (*usus*), but *habitus*.[17] So, virtue can also cause something that is a *habitus*. Or, as Andrea Nannini puts it, "the gifts constitute the effects of virtues: effects understood not as the uses (acts) of the virtues but as the relative habitual dispositions" (Nannini 2022, 44).

To conclude, Langton employs the term *habitus* in the discourse on virtue. He uses this term with different intended meanings, sometimes in a broad sense, like something which we have as ready, or in a very spe-

---

quam proximum', 'diligit' copulat facultatem ibi. Potest dici quod alia est facultas qua diligit deum et alia qua diligit proximum, et una est magis intensa, alia minus, licet ex eadem caritate surgant."

16 Stephen Langton, *QT*, q. 76.2, c. 1.1, p. 160: "Hoc forte impune posset concedi. Set tamen ad presens dicimus quod hoc uerbum 'diligit' in predicta copulat caritatem que unica est in isto. Nec tamen illa notatur esse magis uel minus intensa, quia non notatur excessus respectu caritatis, set respectu effectus eius, scilicet motus; non dico motuum quos habeat, cum dormiens nullos habeat, set motuum in habitu quorum caritas in tali statu est effectiua, sicut cum dicitur "deus magis diligit istum quam illum", non notatur excessus in caritate, quia illa est unica, scilicet diuina essentia, set notatur excessus effectuum quos habebit uel habet in istis.

17 Stephen Langton, *QT*, q. 94, c. 2, p. 372: "dona naturaliter sunt posteriora uirtutibus, quia sunt effectus uel quasi effectus uirtutum – non dico 'effectus', idest usus, set 'effectus', idest habitus."

cial sense, like in q. 76.2, where he speaks about the effect of loving in some person and calls it: *in habitu*. He applies to a problem of virtues the distinction *in habitu* vs. *in usu*., which, according to Colish, comes from Peter Lombard and which was present in Simon of Tournay. Langton also employs the distinction between disposition and *habitus* as transmitted by Abelard. In my view, Langton improves it by stressing that a disposition and a *habitus* that comes from it are in fact the same thing.

6. Thomas Aquinas

Thomas Aquinas introduces a significant change in thinking about virtues, dispositions and *habitus*. Before we start analysing his utterances, let us quote a caveat by Nicholas Austin, which was the main reason to use the term *habitus* in the Latin version in this article:

> There is an acute danger of misinterpreting Aquinas's virtue theory as we read his understanding of habit through modern eyes. Even scholars who recognize some differences between modern and medieval accounts of habit are often unaware of how wide the gap really is. To simply substitute the modern idea of habit for Aquinas's *habitus* would be a significant failure of exegetical accuracy and, more important, would obscure the riches of Aquinas's helpful account. (...) The modern concept of habit therefore picks out a mechanistic pattern of response, generated by multiple repetitions of identical actions. Habits, therefore, lead to unthinking and nonvoluntary action. (Austin 2017, p. 23–24)

Austin refers to Aquinas's theory. However, as we see, this warning applies to all medieval authors. For all of them, *habitus* is not something automatic and not controlled by will. Moreover, it can also mean other different things, if we adopt a broader sense. Anyhow, as we will see, this connection to rational action, which as such is combined with human will, is particularly emphasised by Thomas Aquinas.

In his *Summa theologiae*, Aquinas discusses the problems related to *habitus* and virtues in questions 49–67 of the first part of the second part (Ia-IIae). In q. 49, he considers *habitus* in general as to its substance: whether it is a quality, whether it is a distinct species of quality, whether *habitus* implies an order to an act, the necessity of *habitus*. In the next question, he concentrates on the subject of *habitus*, e.g., whether there is a *habitus* in the body, or whether the soul is the subject of *habitus*, etc. From q. 55, Aquinas starts his analyses on virtues and considers: whether human virtue is a *habitus*, whether it is an operative *habitus*, whether it is a good *habitus*, the definition of virtue.

In article 1 of q. 49 of Ia-IIae, we find the following argument against the starting hypothesis, which contains a definition of *habitus*:

> Further, *every habit (habitus) is a disposition*, as is stated in the Book of the Predicaments (*Categor*. vi). Now disposition is *the order of that which has parts*, as stated in *Metaph*. v, text. 24. But this belongs to the predicament *Position*. Therefore habit is not a quality.[18]

For the first time, we see that *habitus* has been called a disposition (*dispositio*). Moreover, disposition has been clearly defined here, as an order of that which has parts. What is more, both definitions are based on Aristotle's texts. Next, the conclusion of the argument is refuted by the third definition: "On the contrary, The Philosopher says in the *Book of Predicaments* (Categor. vi) that habit is a quality which is difficult to change." In the main reply, Thomas explains that if something "has a relation in regard to itself or to something else" *habitus* is a quality, "since this mode of having is in respect of some quality." In this way, he defends the traditional approach that we already know from previously presented medieval thinkers.

However, is Aquinas right? Does Aristotle really call *habitus* a disposition? In Book V of Metaphysics, the Stagirite writes:

> Having (ἕξις) means (a) in one sense an activity, as it were, of the haver and the thing had, or as in the case of an action or motion; for when one thing makes and another is made, there is between them an act of making. In this way between the man who has a garment and the garment which is had, there is a "having." Clearly, then, it is impossible to have a "having" in this sense; for there will be an infinite series if we can have the having of what we have. But (b) there is another sense of "having" which means a disposition (διάθεσις), in virtue of which the thing which is disposed is disposed (διάκειται τὸ διακείμενον) well or badly, and either independently or in relation to something else. E.g., health is a state, since it is a disposition of the kind described. Further, any part of such a disposition is called a state; and hence the excellence of the parts is a kind of state. (Aristotle, *Metaphysics*, V, 1022b, transl. Tredennick)

This is exactly the fragment that Aquinas refers to. We clearly see that Aristotle indeed affirms that *habitus* is a disposition. In the above quote, I have added key Greek words from the original text to emphasise that

---

18 This and the next quote literally translated by the Fathers of the English Dominican Province (Online Edition 2017 by Kevin Knight, New Advent). See original wording: Thomas Aquinas, *Summa theologiae*, Ia-IIae, q. 49, a. 1: "Preterea, omnis habitus est dispositio, ut dicitur in praedicamentis. Sed dispositio est ordo habentis partes, ut dicitur in V Metaphys. Hoc autem pertinet ad praedicamentum situs. Ergo habitus non est qualitas."

Aristotle really speaks about ἕξις, consequently translated to Latin as *habitus*, and that he really defines it as διάθεσις, which according to leading Greek-English dictionaries, such as DMLBS, means exactly: 'a disposition, arrangement.'

So, Aquinas is perfectly right when quoting Aristotle. What is more, we see that the approach that distinguishes *habitus* and disposition, presented by Peter Abelard, Alan of Lille, Simon of Tournai and to some extent by Stephen Langton, is not coherent with Aristotle's concepts of ἕξις and διάθεσις. Aquinas, who reads original works of the Stagirite, becomes more faithful to his theories and finally presents *habitus* as disposition.

Now, the last step is to answer the question: how does Thomas define virtue? In article 1 of q. 55 of Ia-IIae, in the main answer, he states that "virtue denotes a certain perfection of a power (*potentia*)." A few sentences later, he adds:

> Now there are some powers which of themselves are determinate to their acts; for instance, the active natural powers. And therefore these natural powers are in themselves called virtues. But the rational powers, which are proper to man, are not determinate to one particular action, but are inclined indifferently to many: and they are determinate to acts by means of habits, as is clear from what we have said above (q. 49, a. 4). Therefore human virtues are habits.[19]

Hence, rational powers which are determinate to their acts, which are called virtues, do it by means of *habitus*. In the light of this quote, we can add that it is because of the disposition, which makes actions ordered, and thus, directs them or determines them towards the proper end. So, because it is being done by means of *habitus*, Aquinas concludes, virtue is *habitus*. But habitus is a disposition. Therefore, according to Thomas Aquinas: virtue is disposition.

7. Conclusions

The analyses provided by the authors mentioned in this study show that disposition was differently conceptualized by medieval analytic authors. The notion of disposition has not always fit with the concept of virtue.

---

19 Ibid., q. 55, a. 1, co.: "Sunt autem quaedam potentiae quae secundum seipsas sunt determinatae ad suos actus; sicut potentiae naturales activae. Et ideo huiusmodi potentiae naturales secundum seipsas dicuntur virtutes. Potentiae autem rationales, quae sunt propriae hominis, non sunt determinatae ad unum, sed se habent indeterminate ad multa, determinantur autem ad actus per habitus, sicut ex supradictis patet. Et ideo virtutes humanae habitus sunt."

Although as Nederman claims, philosophers reported by Peter Abelard, initially referred to these two key terms used by Aristotle, namely: ἕξις and διάθεσις, and accepted his statement that ἕξις belongs to the category of quality, they did not follow him in defining ἕξις as διάθεσις. Quite the opposite, they distinguished these two concepts, *habitus* and *dispositio*, and claimed that the first one is difficult to change, whereas the latter is easily changeable. This approach was rooted in Aristotle's utterance presented in *Categories*, 8. However, they neglected his further statement, according to which ἕξις is διάθεσις, and they did not know *Metaphysics*, V, which in their times had not been translated into Latin yet. In consequence, these thinkers and the next generations, represented by Alain of Lille or Simon of Tournai, and later by Stephen Langton, decided that virtue can only be *habitus* and not *dispositio*.

Moreover, as Colish rightly indicated, in twelfth-century thought, *habitus* was not understood only in an Aristotelian way. We have presented clear illustrations which confirm her opinion, at least in Langton's considerations. We have also demonstrated that the term *habitus* played an important role in the discussions concerning virtues, but in a very different context. For instance, it was employed to distinguish between virtues *in usu* and virtues *in habitu*. We have also considered different uses of this term to speak about the ontological status of loving.

However, we have also observed that step by step medieval thinkers were trying to connect these two concepts back. John of Salisbury indicated that a disposition can be transformed into a *habitus*. Hence, it becomes *habitus* when it is unlikely to change. So, virtue, which is *habitus*, is something that was previously a disposition. Next, Stephen Langton pointed out that, in fact, a disposition and a *habitus* which comes from a disposition are one thing.

Finally, Thomas Aquinas went back to Aristotle's texts and faithfully employed his conceptual framework. As a result, he stated that *habitus* is disposition, so, we can say the same about virtue.

To answer the questions posed in the introduction, we can say that according to Aristotle's theory, adopted by Thomas Aquinas, moral virtues, such as courage, temperance, friendliness or justice, are dispositions. Next, we can also confirm that the Greek "hexis" used in the Aristotelian definition of virtue is a kind of disposition. Finally, the Aristotelian approach "virtues" and "dispositions" is not the only possible way of dealing with these terms. There were different approaches which are also plausible and perhaps even valuable.

However, these answers are not the only benefit of analysing medieval texts. In my view, there are important lessons we can learn from this short history of the use of the three terms: "virtue," "disposition," and "*habitus*", in the texts of analytic medieval thinkers. The meaning of such terms themselves is vague. It depends very much on the context or on the conceptual framework in which they are employed. It was enough to tear them away from Aristotle's text to understand them differently. This conclusion refers to both dispositions and *habitus*. Disposition was restricted only to something easily changed. *Habitus* was understood as something different from disposition and also acquired a broader meaning.

We can also notice that it was very easy to adopt a sharp distinction and lose a deep understanding of what a disposition is. It seems that the medievals before Aquinas understood dispositions as a sort of readiness, and did not see that it denotes a special arrangement which enables specific actions. Otherwise, they could see that being a *habitus* does not exclude being a disposition – which, according to one modern translation, is a "settled" one. This deep meaning has been restored by Aquinas thanks to his careful reading of Aristotle, and virtues "became" dispositions again.

References

Aristotle, *The Nicomachean Ethics of Aristotle*, transl. and introd. D. Ross, Oxford University Press, London (1925).

Aristotle, *Nicomachean Ethics*, transl. H. Rackham, William Heinemann Ltd, Harvard University Press, London–Cambridge, Mass. (1956).

Aristotle, *Nicomachean Ethics*, in: *The Complete Works of Aristotle*, vol 2, J. Barnes, Princeton University Press, Princeton, N.J., 1729–1867 (1984).

Austin N., *Aquinas on Virtue: A Causal Reading*, Georgetown University Press (2017).

Azzano, L., Raimondi, A. "Vices, Virtues, and Dispositions." *TheoLogica: An International Journal for Philosophy of Religion and Philosophical Theology*, 7(2) (2023), 87–110. https://doi.org/10.14428/thl.v7i2.67873

Bieniak M., "Faith and the Interconnection of the Virtues in William of Auxerre and Stephen Langton", in: *Fides Virtus: the Virtue of Faith in the Context of the Theological Virtues. IGTM International Conference Proceedings (Padua, 6 to 9 July, 2011)*, ed. R. Quinto, S. Vecchio, Aschendorff, Münster (2014), 209–220.

Colish M.L., "Habitus Revisited: a Reply to Cary Nederman," *Traditio* 48 (1993): 77–92.

Cunningham S.B., "Albertus Magnus and The Problem of Moral Virtue," *Vivarium* 7(2) (1969), 81–119.

Goris, H.J.M.J. "Acquired and Infused Moral Virtues in Wounded Nature," *The Virtuous Life*, eds. H.J.M.J. Goris, H.J.M. Schoot, Peeters, Leuven – Paris – Bristol, CT (2017), 21–46.

Hursthous R., Pettigrove G., "Virtue Ethics," in: *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), eds. E.N. Zalta, U. Nodelman, URL: https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/

Kraut R., "Aristotle's Ethics," in: *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), eds. E.N. Zalta, U. Nodelman, URL: https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/

Laemers, J.W., "Claustrum animae: The Community as Example for Interior Reform," in: *Virtue and Ethics in the Twelfth Century*, eds. István P. Bejczy and Richard G. Newhauser. Leiden, The Netherlands: Brill, Leiden–Boston (2005), 119–130. https://doi.org/10.1163/9789047407270_008

Mattison III W.C., "Aquinas, Custom, and the Coexistence of Infused and Acquired Cardinal Virtues," *Journal of Moral Theology*, 8(2) (2019), 1–24.

Nannini A., "Fear," in: Stephen Langton, *Quaestiones Theologiae, Liber III, Volume 2*, eds. M. Bieniak, M. Trepczyński, W. Wciórka, Oxford University Press, Oxford (2022), 41–61.

Nederman C.J., "Nature, Ethics and the Doctrine of Habitus: Aristotelian Moral Psychology in the Twelfth Century," *Traditio* 45 (1989–90): 87–110.

Sockett H., "Dispositions as Virtues: The Complexity of the Construct," *Journal of Teacher Education* 60(3) (2009), 291–303. https://doi.org/10.1177/0022487109335189

Stephen Langton, *Quaestiones Theologiae, Liber III, Volume 2*, eds. M. Bieniak, M. Trepczyński, W. Wciórka, Oxford University Press, Oxford (2022).

Swanstrom J.L., "Aquinas and Virtue Acquisition in Secondary Causes," *Revista Portuguesa de Filosofia* 76(1) (2020), 261–282.

Torchia N.J. "The Limits of Virtuous Action: Aquinas on the Subjective Conditions of Moral Virtue," *Angelicum* 79 (2002), 855–877.

Whitmore A.C., *Dispositions and Habits in the Work of Saint Thomas Aquinas*. A dissertation, Faculty of the School of Theology and Religious Studies of The Catholic University of America, Washington, DC (2018).

6. DISPOSITIONS AND EMOTIONS

https://doi.org/10.18485/baf.2025.2.ch12

Damir Smiljanić

COOL PHILOSOPHY OR THE ART OF RESTRAINING FROM JUDGING WITHOUT BEING INDIFFERENT

Abstract: Although philosophy is primarily based on the exchange of arguments in support of or against a thesis, it is also based on certain cognitive and value *attitudes* supporting the argumentation. When discussing a problem, it is not the opponents philosophers *care* about but being "right". Were their argumentation not guided by such an attitude the philosophical discussion would turn into an arbitrary sequence of statements without any focus on the very matter. The problem arises in the case of self-reflection – when philosophers realize that they are starting from certain presuppositions that limit their attempt to solve the problem – or meta-reflection – when it becomes clear that reaching a consensus in the discussion is impossible due to the different starting points of its participants. This insight might have a distancing effect with regard to one's assumptions or the assumptions of philosophizing in general. To achieve such (self)distancing, it is necessary to assume a different attitude in the discussion – or to assume none at all! The goal of this presentation is to investigate the *suspension of attitude*, avoiding a one-sided identification with one point of view and enabling a broader perspective on the constellation in which representatives of different positions are involved in the discussion of a problem. In the history of philosophical thought, there have been attempts to formulate such a neutral habitus: Pyrrhonians sought *suspension of judgment* (*epoché*), Adam Smith proposed the instance of an *impartial spectator* during the assessment of morally sensitive situations, Max Weber advocated *value neutrality* when dealing with science. The key question is whether the restraint we are discussing here rests on a special (personal) *disposition* or *value attitude* gradually built up and accepted with full awareness. *Indifference* as a character disposition is not suitable as a basis because it is essentially anti-theoretical. Another basis must be found on which to establish distance from philosophical biases, and it will include *coolness* (in the sense of the absence of strong emotions) and *prudence*. At the same time, however, there will be *engagement* in the matter because without this philosophy loses its seriousness.

Keywords: argumentation, disposition, value attitude, restraint, impartial spectator, value neutrality, Pyrrhonians, Adam Smith, Max Weber.

1. The Initial Problem

First of all, I would like to introduce the problem I will be dealing with in my paper. My starting point is the distinction between *dispositions* and *values*, as indicated in the title of the conference. There is something like *indifference*, a quality we attribute to people who do not develop a particular attitude towards a certain issue or situation. One is indifferent if one does not care whether an issue or a situation will be perceived in a positive or negative light. In other words: one does neither have a positive nor a negative attitude towards the issue or situation at hand, therefore, one does not evaluate it. It could even be said that the presupposition for evaluating something as good or bad is the lack of indifference towards the object of evaluation. Indifference, in this sense, makes the act of biased evaluation impossible or at least difficult. Indifference could, thus, be seen as a *non-evaluative disposition* in contrast to, for example, benevolence as a positively or callousness as a negatively evaluated disposition. However, I am interested in finding out whether indifference can also be recognized as *a disposition for evaluation* in certain cases, that is, as a *manifestation* of a *value attitude*.

In order to understand the meaning of indifference, its *psychological description* should be distinguished from the *philosophical* one. Indifference can be attributed to someone as a *character trait*. We can recognize an indifferent person by the way they[1] behave and act: their reactions are rarely intense, and their bodily movements are rather slow and almost lifeless. Such a person seems to have no feelings, and if they do show them, their expression lacks vitality. From a semantic point of view, this description is related to the *phlegmatic temperament* described in the framework of ancient medicine and psychology. According to the system of *humoral theory*,[2] a person whose physical constitution (dominance of cold body fluids: *phlegm*) correlates with calmness, slowness, and, to some extent, lethargy is described as phlegmatic. Paradoxically, one could say that someone with a phlegmatic temperament appears to be completely *untempered*. In a philosophically (more precisely: ethically) relevant sense, indifference is a characteristic of a person lacking *interest* in something, which means that in a situation where they need to act in accordance with advice or instructions, they are not going to fulfill the task with tremendous vigor and intensity and will rather give the impression of not caring. So, here, indifference is characterized by the readiness to act or perform

1 For stylistic reasons, I will use the singular form of the pronoun "they" instead of masculine and feminine pronouns.

2 On the historical development of this psychological theory cf. Arikha, 2007.

a task, and all of this obviously depends on the *attitude* towards the matter itself. In other words, the trait of indifference refers to *character* in a psychological sense, and to a specific *attitude* in a philosophical or ethical sense. Or, in accordance with the main topic of this conference: indifference in the psychological sense is a *character disposition*, and in the philosophical sense a *value attitude*.

Although the consideration of the character of indifference is primarily of psychological and, seemingly, axiological importance, I would like to deal with this characteristic – in a dual sense: as a disposition and as a value attitude – from the perspective of *metaphilosophy*. Hence, I am interested in the status of indifference in the field of philosophy. Metaphilosophy is a discipline exclusively dealing with the reflection of the structure of philosophical speech and thought, without an explicit attitude towards philosophical pretensions to possessing truth. It is not intended to identify the philosophical position corresponding to reality, but to help understand what it actually means to take up a philosophical position. Does this mean that the metaphilosopher should be *indifferent* to philosophers arguing the case of their thesis in order to defend it from objections of those who disagree? If a metaphilosopher (at least in an ideal case) does not interfere in philosophical discussions and disputes, does this mean that the solution to a problem does not concern them and that they "do not care" which of the participants in the discussion is right? If this is the case, could they be accused of being "indifferent" to or even "ignorant" of philosophy? Or might the fact that they behave this way, i.e. that they remain "cool-headed" when dealing with philosophical discussions, rather be an indication of some kind of excellence, the ability to preserve impartiality in the face of philosophers' *non-objectivity*? These issues should be considered separately, so I will skip them here. Regardless of the answers to these questions, I think that metaphilosophy can serve as a useful aid to orientation in the field of philosophy, even if it does not evaluate the pretensions of philosophical authors concerning the truthfulness (or at least truthlikeness) of their theories.

Now I would like to point out the tendencies of gaining control over situations in order to avoid the adoption of a dogmatic attitude in discussions that are abundant in philosophy itself. The goal is to show that it is not only possible "outside" of philosophy, i.e. on a metaphilosophical level, to distance oneself from a standpoint and simultaneously demonstrate one's engagement in thinking but also in the framework of philosophy itself. To be indifferent in philosophy may also be a sign of wisdom and a prerequisite for gaining wisdom.

2. How to Restrain from Judging in a Philosophical Manner

My intention is to illustrate with a few examples how certain thinkers restrain from judgment within philosophy. I will consider whether this is done more passively, based on some disposition of the character, or rather actively, by building a position on certain arguments. In other words, we will see whether someone is predisposed to not adhering to a specific philosophical position due to their character or whether they are able to distance themselves from certain philosophical solutions to problems due to a belief reached through critical self-reflection. The three examples I will consider can also be understood as forms or ways of restraining from philosophical reasoning: 1) restraining from judgment in the ancient tradition of *radical scepticism* (*Pyrrhonism*) (the so-called *epoché*), 2) observing a problematic situation through the eyes of an unbiased observer (as suggested by Adam Smith), 3) Max Weber's recommendation that those dealing with science should – at least at the academic level – be *unprejudiced*. Let us consider these examples one by one.

2.1. Suspension of Judgement (*epoché*)

In some phases of the history of philosophy, a certain exhaustion from fruitless discussions is noticeable, when one party to the dispute does not accept the standpoint of the other and vice versa. It is, of course, possible to introduce new perspectives on the matter discussed. This, however, only multiplies points of contention and further postpones the possibility of resolving the dispute, be it only by a temporary consensus of all parties involved. Apart from that, there is a radically different strategy striving to avoid rather than resolve the dispute – namely, excluding philosophers (if they can even be called that) completely from such discussions. This alternative approach to philosophy, which is, in fact, rather some kind of abandonment of philosophy or distancing from it, should have been an alternative to the *dogmatic attitude* and *academic scepticism* in late antiquity. Adherents to Pyrrhon's philosophy, the so-called *Pyrrhonists*, renewed his teachings at the beginning of the new millennium. Sextus Empiricus, who was actually a physician, systematized this teaching in *Outlines of Pyrrhonism*. It is the most important document for the study of *radical scepticism*. The main idea of this sort of philosophizing is outlined in the following quote:

> Scepticism is an ability to set out oppositions among things which appear and are thought of in any way at all, an ability by which, because of the equipollence in the opposed objects and accounts, we come first to suspension of judgement and afterwards to tranquillity.[3]

The concept of *equipollence* is defined as follows:

> By 'equipollence' we mean equality with regard to being convincing or unconvincing: none of the conflicting accounts takes precedence over any other as being more convincing.[4]

Ultimately, the main goal of radical scepticism is to gain *tranquillity* of the soul by *suspending* the judgement on philosophical problems:

> Suspension of judgement is a standstill of the intellect, owing to which we neither reject nor posit anything. Tranquillity is freedom from disturbance or calmness of soul.[5]

What is the point of this sceptical strategy? In philosophical discussions, dogmatic participants usually opt for one thesis and reject the counterthesis, i.e. they bring forth good arguments in support of the former or argue against the latter. When discussing with their opponents, they are not indifferent at all – on the contrary, they are highly engaged, frequently presenting their arguments or objections with emotional involvement. This makes philosophical discussions to a certain extent intense and fierce. Pyrrhonian sceptics, on the other hand, deem restraint from final judgement and remaining undecided, i.e. not opting for any position, the best approach, as they realize that every relevant statement in philosophy can be supported by an argument, but also disputed by a counter-argument. The reason they deem this the best approach is psychological – namely, the recognition that indifference towards the philosophical discussion of an issue leads to *tranquillity*, the ideal ancient thinkers aspired to. In other words, the suspension of philosophical reasoning helps to achieve tranquillity. The reason why sceptics stay out of philosophical discussions is simply pragmatic: they just do not want to be upset by discussions. Their primary goal is to stay calm and find inner peace.

Does this mean that Pyrrhonians are *indifferent* to philosophy? It seems that they are not interested in which of the philosophical positions is "right", which of the philosophical statements "corresponds" to reality

---

3 Sextus Empiricus, 2000, p. 4.

4 Op. cit., p. 5. In other words, all philosophical perspectives are equally right or wrong. (Therefore, in Smiljanić 2006, the author demonstrates the importance of *perspectivism* for metaphilosophical discourse.)

5 Ibid.

and what reality is "in itself". Dogmatic philosophers would denote sceptics "anti-philosophers" (Plato would probably consider their position "mysological"). However, it would be too hasty or even dogmatically malicious to consider radical sceptics the opposite of philosophers. They *care* about philosophy as well, explicitly identifying themselves as philosophers and comparing themselves to dogmatic thinkers and academic sceptics. Moreover, their radical criticism of all preceding philosophy (including their own!) is due to their striving for correct opinions and developing well-thought-out attitudes that influence everyday practice. However, criticism in this context does not represent the mere destruction of a position, but rather a suggestion for autocorrection with *psychotherapeutic* value – a kind of *mental therapy* for philosophers. Although they did not yet have this intention, we could say that the Pyrrhonians' way of speaking and thinking laid the foundations for a *psychology of philosophy*: the task of this approach would be to examine why proponents of a certain thesis stick to their position or even defend it rather aggressively (e.g. with a certain anger towards those who do not share their point of view), when consensus with other thinkers is lacking or adequate solutions to a problem cannot be reached. The reasons for such behavior cannot be of a purely theoretical or even rational character, as the minimum rational condition for a successful discussion is, first and foremost, to listen to our opponent without interrupting him or her and drop any negative emotions during the discussion. Philosophical discourse should not be a training ground for inappropriate emotions and resentments. It should rather provide an opportunity for open discussion and fair exchange of arguments, even if the viewpoints of the discussants diverge diametrically. This might be an idealization of this kind of discourse, but at least it is an ideal one should strive for in order to avoid dogmatism and intellectual narcissism. After all, dogmatism has a negative connotation regardless of this philosophical context. The radical sceptics can take the credit for pointing out the necessity of critical self-reflection concerning one's own point of view and were thus – at least I assume – responsible for developing *metaphilosophy* as a special form of thinking.[6]

6 "Scepticism is a kind of liberation philosophy or, perhaps better, a liberation *from* philosophy. It frees us from involvement with all those problematic issues and lifts from our shoulders the burden of having to take sides in all those abstruse controversies." (Rescher, 1985, p. 248) Such a position is akin to that of metaphilosophy. However, unlike the latter, scepticism advocates the thesis that philosophy is a research project doomed to fail, because no philosophical statement can be fully proven true and defended against other thinkers' objections. In other words, scepticism itself is a philosophical position, albeit with a negative connotation. Cf. Rescher's defense of

2.2. The Impartial Spectator

The way prejudices or affects hinder or even prevent the perception of all aspects of a situation relevant to making certain judgment or decision is especially thematized in *ethical* contexts. Whether someone acted in a morally correct manner depends on whether they were bound by their own domain of interest or whether they took into account the interests of others as well. In order to recognize the interests of others (at least to a certain extent), one must be able to put oneself in their position and, if necessary, account for their feelings as influential factors in human actions. We can call this process of imagining another's situation *sympathy*. In modern British moral philosophy, the term 'sympathy' is used to indicate the ability of a person to put themself in another's position, at least in their mind. In his book *Theory of Moral Sentiments* (1759) one of the main representatives of that philosophy, Adam Smith, considers the concept of sympathy to be crucial for understanding how we acquire the competence of moral reasoning in relation to our own as well as other people's actions and behavior. We cannot directly feel what a person is experiencing, but we can sympathize with them thanks to our ability to *imagine* what we ourselves would feel in that situation. Smith argues psychologically:

> By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him, and thence form some idea of his sensations, and even feel something which, though weaker in degree, is not altogether unlike them.[7]

The main difficulty in (imaginatively) taking another person's point of view is the strength of our *affects* and *feelings* coupled with our *self-love*:

> When we are about to act, the eagerness of passion will seldom allow us to consider what we are doing, with the candour of an indifferent person. The violent emotions which at that time agitate us, discolour our views of things; even when we are endeavouring to place ourselves in the situation of another, and to regard the objects that interest us in the light in which they will naturally appear to him, the fury of our own passions constantly calls us back to our own place, where every thing appears magnified and misrepresented by self-love.[8]

An important factor in establishing sympathy is *temporal distance* – the more time has passed since we performed an action, the easier it is for

---

philosophical validity claims against their sceptical re-examinations in: op. cit., pp. 248–252.

7 Smith, 2004, p. 12.

8 Op. cit., p. 182.

us to observe its effects from the perspective of an *indifferent spectator*. In Smith's words:

> When the action is over, indeed, and the passions which prompted it have subsided, we can enter more coolly into the sentiments of the indifferent spectator. What before interested us is now become almost as indifferent to us as it always was to him, and we can now examine our own conduct with his candour and impartiality.[9]

What does this mean? At the moment of performing an action, we are more strongly influenced by affects and feelings than afterwards. Therefore, it is much more difficult to make decisions under pressure of a current situation than to judge past events. This also means the affective background makes reasoning in specific situations difficult, and it is hard to achieve a high degree of impartiality under such circumstances. Smith's remark almost sounds like a proverb:

> The man of to-day is no longer agitated by the same passions which distracted the man of yesterday,[10]

meaning that temporal distance is a prerequisite for appropriately judging a situation. (The use of 'cold' is also interesting: the less strong feelings influence our judgment, the "cooler" we become as impartial observers.) Furthermore, Smith identifies people's inability to move beyond self-interest to take up the position of an impartial observer as the main cause of their problems. Thus, he also calls self-deceit a "fatal weakness of mankind".[11] That is why it is necessary to take a step back and observe ourselves "through the eyes of others":

> If we saw ourselves in the light in which others see us, or in which they would see us if they knew all, a reformation would generally be unavoidable.[12]

We can say that the attitude of indifference is embodied in Smith's idea of an impartial spectator: this observer is defined by the absence of interest in a certain situation in which an adequate judgment or decision needs to be made. Smith's (optimistic) assumption is that every individual is capable of activating that instance in themself, and can, thus, distance themself from the interests and values they usually adhere to. Consequently, we are not dealing with a mere innate disposition, but rather with the cognitive ability of evaluation, the success of which increases

9 Ibid., p. 183.
10 Cf. ibid., p. 183.
11 Cf. ibid., p. 184.
12 Ibid.

with each common situation providing an opportunity to practice. It is, therefore, a learning process leading to the formation of an attitude and does not just happen on its own. The subject's activity is required, while the disposition of an indifferent person is usually of a passive character. If, on the other hand, we were to compare Smith's ethical principle with the teachings of Pyrrhonian sceptics, we could say that an impartial spectator would be acceptable to sceptics, because, thanks to taking the perspective of such a spectator, it would be possible to compare two or more philosophical theses to find equally good arguments in favor of each of them and thus establish equivalence between them (*isostheneia*). Both Smith and the Pyrrhonians describe a way to adopt and maintain an indifferent attitude towards things, judgments and situations, but in a way that is in accordance with rational behavior and moral norms.[13] Their common intention is to avoid dogmatism and the influence of affects and passions on the process of correct reasoning. This should be kept in mind when I speak about the "coolness" of philosophy in accomplishing its intellectual mission in the paper's final section.

2.3. Value Neutrality

The third example of how to avoid bias in judgements does not come from a purely philosophical, but rather from a sociological perspective (however, related to the cultural-philosophical one). In fact, there is a relation to *ethical* considerations, but the focus shifts to the value orientation of those professionally dealing with science: *academic teachers*. Of course, we are talking about the famous lecture of the German sociologist Max Weber "Science as a Vocation" [*Wissenschaft als Beruf*], held in 1917, addressed to students in Munich and published in 1919. Weber's text is equally interesting for the sociology of education as for pedagogical philosophy and ethics of science. It is a plea for the *value neutrality* of science, namely an appeal to lecturers from various disciplines – especially those from the domain of humanities – to strictly avoid imposing their political and other value-driven attitudes, and to prevent them the indoctrination of students by suggesting that certain social values or political orientations are desirable or not. Weber thus advocates the idea of value neutrality, which means insisting on a strict separation between purely scientific and political or other socially relevant interests. Without this, he argues, scientific activity would turn into mere agitation and universities would become arenas for politico-ideological struggle. Weber explicitly rejects the politicization and ideologization of science:

13 The Pyrrhonian sceptics repeatedly draw attention to the fact that their behavior is compatible with "laws and customs". Cf. Sextus Empiricus, 2000, pp. 7, 9.

> It is often said, and I subscribe to this view, that politics has no place in the lecture room. [...] [T]he genuine teacher will take good care not to use his position at the lectern to promote any particular point of view, whether explicitly or by suggestion. For this latter tactic is, of course, the most treacherous approach when it is done in the guise of "allowing the facts to speak for themselves."[14]

Value judgment prevents or hinders orientation towards the facts on which scientific research should be based:

> I am willing to demonstrate from the writings of our historians that whenever an academic introduces his own value judgment, a complete understanding of the facts *comes to an end*.[15]

Weber's plea for freedom from value judgments should not be understood in the sense that science itself should be purified of any interest. That would be self-contradictory because every science comes at least with a research interest: a mathematician is interested in studying the relationships between numbers, a physicist in describing the structure of matter, a biologist in the functioning of living beings, a historian in interpreting the cause of historical events and the motives of the people involved, a sociologist in gaining insight into how a social system is organized, etc. This, however, does not just denote any interest, but a *scientific* one – specifically, the interest that leads to knowledge. Therefore, it is not surprising that Weber points out with a certain pathos:

> [I]n the realm of science, the only person to have "personality" is the one who is *wholly devoted to his subject*. And this is true not just of science.[16]

Consequently, science is not devoid of interest; rather, it is based on commitment to a certain issue. Indifference towards the subject matter itself is impossible, but the crucial point is that a person engaged in science needs to distance themself from particular interests and value preferences while engaged in science. A scientist can prefer a political or ideological position only outside of science, and never propagate their values *ex cathedra*. In short (expressed in Weber's words): science needs teachers, not leaders. From Weber's considerations on the "vocation" of science, one could therefore draw the philosophically relevant conclusion that there are two types of "indifference": one towards the matter itself, where some non-scientific interest usurps the matter, and one towards values, in this case towards those who are externally threatening the neutrality of sci-

14 Weber, 2004, pp. 19–20.
15 Op. cit., p. 21.
16 Ibid., p. 10.

entific research. A person who is indifferent cannot be a good scientist, while someone being indifferent to non-scientific values ensures the objectivity of their research. Although Max Weber's consideration of this issue arises from sociological and pedagogical concerns, the conclusion remains useful for my final comment on the distinction between dispositions and values using the example of indifference.

3.3. Conclusion: How to Keep a Cool Head in Philosophy

Following this review of some examples from the history of philosophy and ethics (of science), we return to the starting point of this presentation: Is indifference merely a disposition for a certain behavior or something more – possibly an expression of a certain value attitude? Furthermore, how should we evaluate indifference in philosophy? Is it to be desired or to be avoided? I think the examples discussed – the sceptics' radical suspension of judgment, Smith's concept of the impartial spectator, and Weber's appeal for value neutrality in science – demonstrate that, in a philosophical context, indifference is indeed a positive phenomenon. Indeed, we could say that it constitutes an ethically desirable habitus. So, indifference holds a certain value, at least within philosophy. We could even say that an indifferent attitude is a necessary condition for "effective" philosophizing. This can be confirmed by all three examples of an indifferent attitude discussed in my paper.

Let us start with the suspension of philosophical judgment. According to Sextus Empiricus and other proponents of Pyrrhonian scepticism, a "real" philosopher is expected to restrain from judgement in philosophical disputes, given that it is mostly representatives of dogmatic positions that are involved in these.[17] To avoid unproductive discussions with dogmatic philosophers, it is wisest not to advocate any position regarding the content of a philosophical question (e.g. offer a special solution to a philosophical problem), but to refrain from judgment and simply stand aside. What is key to this view of practicing (some may say: *not* practicing) philosophy is that we achieve tranquillity – Pyrrhonian scepticism is actually a kind of philosophical therapy: a remedy against excessive pretensions and claims to the exclusive right to the truth. It is regrettable that the *metaphilosophical* importance of this specific philosophical orientation is not sufficiently recognized.

17 *Academic sceptics* stand between dogmatism and radical scepticism and their agnosticism implicitly presupposes at least the possibility of the (ultimate) knowledge of "things in themselves", which is inconsistent and self-contradictory, and may even be a reason for a new dogmatic attitude. For more information about the difference between academic and radical scepticism see Sextus Empiricus, 2000, pp. 57–62.

When it comes to Smith's conception of the impartial observer, it could serve as a kind of *ideal* for philosophical thinking: in every discussion, especially with their dissenters, philosophers should strive to *better understand* not only their own point of view, but also others' perspectives. This can only be achieved if we distance ourselves from our own views and interests as well as from the more or less expressed emotions binding us to our position and if we take up the position of an external observer who is able to reach a more adequate and fair judgment both concerning the matter itself and concerning other perspectives on it. Therefore, the *hermeneutical* relevance of Smith's principle for philosophy consists in better understanding the intentions and motives of other philosophers and being able to appreciate the possible significance of their insights for the solution of a problem. It is *metaphilosophy* that could profit from this principle, as its task is to describe the structure of philosophical discourse and thought as *neutrally* as possible. To achieve this, it is necessary to adopt a position fundamentally distinct from the philosophical one – operating at a *higher level* of reflection, with sufficient distance from philosophy itself. What else could a metaphilosopher be but an *impartial observer of philosophy*!

Certainly, philosophers should also accept Max Weber's recommendation to devote themselves to the study of their subject without evaluation and ideological assumptions, especially if they are teaching at an academic institution. In philosophy, critical thinking alone is not enough to uncover the ideological presuppositions and dubious values guiding our opponents' positions. Philosophical thinking must also be *self-critical* – it requires philosophers to recognize the ideological background of their own thinking. On the other hand, as a connoisseur of Neo-Kantian philosophy, Weber was aware of the need to *distinguish values from facts* and to investigate values as a separate sphere (not of factual being, but of *validity*[18]) in the context of philosophy. (It is one thing to think about values, and another to represent values and submit our thinking and actions to them – the purpose of value neutrality is to prevent the latter from happening.) Last year, at the *First Analytical Forum*, I pointed out Rescher's distinction between *descriptive* and *prescriptive metaphilosophy*, considering that even at this level of thinking one should refrain from strong value judgments about philosophy.[19] I tried to defend the idea of a metaphilosophy that strives to describe and explain philosophical phenomena instead of evaluating them (e.g. as 'true' or 'false', 'good' or 'bad', 'useful' or 'useless' etc.). Therefore, I consider this paper on the character of value indifference to be a contribution to descriptive metaphilosophy.

18 In German: *Geltung*.
19 See Smiljanić, 2024.

Finally, a terminological note: in everyday speech – at least in Anglo-Saxon usage – we describe someone who behaves moderately and thoughtfully in various situations, without strong passion and emotions yet in an acceptable and even desirable manner, as *cool*, since this behavior prevents them from losing composure while keeping the situation under control. A cool person does not allow their own emotions, unexpected circumstances, or other people to "throw them off track". Whatever happens, they stick to their initial attitude, do not change it, and are yet able to overcome any unpleasant situation. Apart from that, we describe someone or something as cool when we appreciate their or its unobtrusive, relaxed quality. This combination of moderation and likability is characteristic of the state of someone or something cool (*coolness*). Semantically, the term 'cool' can be understood in two ways: first, as a reference to physical coldness or our subjective reaction to it; secondly, as an expression for something likable or pleasant. We can say that it simultaneously denotes both the sensory (a certain physically measurable temperature) and the non-sensory (an aesthetically sensitive effect). Although the predicate 'cool' may appear similar to 'cold', they are fundamentally distinct: coolness is a trait that attracts us, as opposed to coldness which typically usually repels (e.g. when we say that we dislike winter's chill). Coolness thus represents a strange kind of coldness – one that irresistibly attracts us.

Hence, we can confidently say that philosophy itself becomes cool when it resists making hasty judgments about intensely debated matters. This philosophical restraint does not stem from mere hesitancy as a sign of character weakness – a mental disposition – but rather reflects a sophisticated position grounded in impartial observation. It demands both witholding judgment and avoiding attachment to particular values when examining complex issues. It is precisely this temperance that makes the mentioned philosopher cool. The fact that we are considering value neutrality as a kind of value attitude does not mean that we are contradicting ourselves, risking a paradox, as if we were to say: "Our interest is not to be interested in any of the parties in the dispute" or "We opt for the option not to opt for any position". In a Weberian way, when I am devoted to the subject, I can never approach it with prejudice. I rather make room for further consideration and the discovery of new views on the matter. The Pyrrhonists simply said: "We are still investigating."[20] And that is exactly what is cool about philosophy – that its search for truth keeps going on and that by discovering new aspects of reality we enrich our intellectual life.

20 Cf. Sextus Empiricus, 2000, p. 3.

References

Arikha, N. (2007). *Passions and Tempers*: A History of the Humours. New York: HarperCollins Publishers.

Rescher, N. (1985). The Strife of Systems: The Grounds and Implications of Philosophical Diversity. Pittsburgh: University of Pittsburgh Press.

Sextus Empiricus (2000). *Outlines of Scepticism*. Edited by J. Annas and J. Barnes. Cambridge: Cambridge University Press.

Smiljanić, D. (2006). Philosophische Positionalität im Lichte des Perspektivismus. Ein metaphilosophischer Versuch. Marburg: Tectum.

Smiljanić, D. (2024). "On the Distinction between Descriptive and Prescriptive Metaphilosophy". In: *Balkan Analytic Forum. Normativity & Normativity of Art*. Conference Proceedings 19–29. X 2023. Belgrade, Serbia. Edited by M. Trajkovski and E. C. McWilliams. Belgrade: University of Belgrade – Faculty of Philosophy, pp. 35–49.

Smith, A. (2004). *Theory of Moral Sentiments*. Edited by K. Haakonssen. Cambridge: Cambridge University Press.

Weber, M. (2004). *The Vocation Lectures*. Edited and with an Introduction by D. Owen and T. B. Strong. Indianapolis/Cambridge: Hackett Publishing Company.

https://doi.org/10.18485/baf.2025.2.ch13

Isidora Novaković

DISPOSITIONS TO TRAGIC EMOTIONS

Abstract: This paper aims to illuminate whether there are dispositions to tragic emotions and what evokes them. Our focus will be on another approach to the question, namely, if there is something in the tragedy itself that makes us feel certain emotions. We will also investigate various types of scenarios characteristic of Greek tragedy. In order to achieve this goal, we will focus on Aristotle's understanding of tragic suffering from *The Poetics*, as well as his understanding of tragic emotions (fear and pity) from *The Art of Rhetoric*. Our starting point will be the analysis of Aristotle's definition of tragedy. The key components of the definition in our investigation of the main question will be the dramatis personae in action and the evocation of fear and pity. Secondly, we will pay attention to Aristotle's examination of the changes of the tragic hero's fortune. In order to understand this change of fortune we will explore *hamartia*, a fatal flaw which leads to the downfall of the protagonist. After this analysis, we will return to the definition of tragedy. And with the tragedy's components and ends taken into consideration, we will ultimately arrive at dispositions to tragic emotions and their significance in clarifying Aristotle's approach to tragedy. The particular examples from the Greek tragedies such as Aeschylus' *Agamemnon*, Sophocles' *Antigone* and *Oedipus Rex* and Euripides' *Iphigenia in Aulis* will greatly help in our investigation. By contrast, we can look at Euripides' atypical tragedy *Medea*.

Key words: *dispositions, fear, pity, hamartia, tragedy*

1. Introduction

We say that a glass has a disposition to shatter even when it doesn't break. Objects have dispositions which do not necessarily manifest in their behavior. Still, these dispositions will manifest if certain conditions are met. Here, we will explore what tragic emotions could be and whether there is a way for them to emerge. The question of dispositions to tragic emotions can be approached from various angles. We can investigate the characteristics of subjects and recipients, prone to react to tragedies in certain ways. Yet, our primary interest is in approaching this question from a different angle, by exploring whether there is something in the tragedy itself that makes us feel that which we will call tragic emotions. Our quest will be greatly helped by the tragic scenarios typical of Greek tragedy, but the crucial line of thought explored will be Aristotle's.

This paper will propel us to look into Aristotle's understanding of emotions as a crucial element of his conception of moral virtues. That is because virtue is formed from the alignment and the constant regulation of emotions. The reasoning behind this inquisition into virtue is to properly understand the tragic emotions that are being evoked in us by the protagonist's actions. Lastly, we will arrive at an understanding of the function of tragedy as catharsis, attained by feeling fear and pity. All of this will be illustrated by particular examples from the selected Greek tragedies.

Showing that we can speak of dispositions to tragic emotions in the aforementioned sense could also have practical value. Namely, applying the dispositional conceptual framework to Aristotle's investigation of tragedy could help refine literary criticism. If a tragedy seems incapable of evoking fear and pity that lead to catharsis in its audience, we can safely assume that it is not fulfilling its purpose. Thus, all of this will show that tragedy's end is the manifestation of its dispositions to evoke tragic emotions.

2. Tragedy

Aristotle defines tragedy as

> an imintation of an action that is admirable, complete and possesses magnitude; in language made pleasurable, each of its species separated in different parts; performed by actors, not through narration; effecting through pity and fear the purification of such emotions. (Aristotle, 2004, p. 244)

As a man must possess, Aristotle explains, moral characteristics, and tragedy portrays people in action, and has a goal to present us with characters on stage, people who can be better, worse or similar to us – those characters themselves will have certain virtues and flaws. Below, for the sake of simplicity and easier understanding, we will designate these people in action as dramatis personae. Dramatis personae will be placed in scenarios, because, Aristotle asserts, "[s]ince the imitation is performed by actors, it follows first that the management of the *spectacle* must be a component part of tragedy." (Aristotle, 2004, p. 244)

The second key component of the definition relevant to the start of our inquiry concerns the nature of the plot of the tragedy. We will be interested in the "events that evoke fear and pity. These effects occur above all when things come about contrary to expectation but because of one another." (Aristotle, 2004, p. 249) Hence, if tragedy is to accomplish its goal – to evoke tragic emotions in us – its plot has to surprise us. But, that unexpected situation still has to be justified and caused by previous events, and not pulled out of the hat like a rabbit. Aristotle especially emphasizes the importance of causation within the different aspects of a tragedy and advises that

> [the poet] ought always to seek what is either necessary or probable, so that it is either necessary or probable that a person of such-and-such a sort say or do things of the same sort, and it is either necessary or probable that this [incident] happen after that one. (Aristot. *Poet.* 1454a34–8)

2.1. Dramatis personae in action

Considering the sheer number of places where Aristotle emphasizes this idea, we can say that the essence and the most elementary structure of a tragedy is its plot, "an imitation of an action, and the action is performed by certain agents." (Aristotle, 2004, p. 244) This would mean that, even if we were to discard all other elements of a tragedy, it could be successful as long as the dramatis personae provide a dramatization of the story. For that we need a dramatization of the events, and not mere reliance on words to tell the story.

Daniels and Scully interpret tragedy's imitative characteristics following Walton's concept of *mimesis* as a game of make-believe (Daniels, Scully, 1992, p. 206). Currie tells us that in Walton's theory "representational works – pictures, plays, novels, films – are props in games of make-believe or imagination" (Currie, 1993, p. 367). Aristotle himself emphasizes tragedy as an art that represents through imagination (Aristotel, 2015, p. 57), and views imitation itself as something that "comes naturally to human beings from childhood" (Aristotle, 2004, p. 241). In that vein, Daniels and Scully view tragedy as precisely an imitation of the plot, which is built from connected events. This plot is enacted by the actors portraying dramatis personae through make-believe (Daniels, Scully, 1992, p. 206). Following this idea, we can picture ourselves in the tragic hero's position, and with that, as Nehamas notes, we can imagine how we would feel in that position (Nehamas, 1994, p. 272).

The events that the tragedy represents, as Daniels and Scully note, can be simple or complex precisely because the real-world events are also like that. As tragedy imitates life, its plot is simple or complex depending on the actions it depicts. Aristotle also makes this distinction and asserts that (Daniels, Scully, 1992, p. 206):

> Some plots are simple, others complex, since the actions of which the plots are imitations are themselves also of these two kinds. By a *simple* action I mean one which is ... continuous and unified, and in which the change of fortune comes about without reversal or recognition. By *complex*, I mean one in which the change of fortune involves reversal or recognition or both. (Aristotle, 2004, p. 249)

Schaurot indicates that the oldest preserved Greek tragedies are amazingly simple (Schauroth, 1932, p. 353).

And truly, if we look, for instance, at Aeschylus' Agamemnon, we see a tragedy set in the broader context of the Trojan War, but its plot focuses only on a fragment of time shared by all the dramatis personae. Namely, Agamemnon returns to his palace with the daughter of the Trojan king (Eshil, 1990, pp. 67–9). They are welcomed by Agamemnon's unfaithful wife Clytemnestra who kills them (Eshil, 1990, pp. 74–5). She feels no remorse for her deed, but proudly pronounces that she killed her husband as revenge for the sacrifice of their daughter Iphigenia (Eshil, 1990, pp. 76–7), his infidelity (Eshil, 1990, pp. 75) and for the sake of ruling the court with her own lover (Eshil, 1990, p. 78). The events of this tragedy are connected and lined up one after the other. As it has no reversals, accidental events or new, unexpected information that leads up to knowledge and, with it, to recognition and/or reversal, this tragedy is simple. This tragedy imitates the action through Clytemnestra, who shows, rather than tells. Although she uses words to describe why she did what she did, her reasoning is neither hidden nor kept secret. Even though we do not see the final blow Clytemnestra delivers to her husband, the sequence of the events in the tragedy ensures that we know that the deed is done, so its later retelling by the Chorus serves not to reveal the event, but to interpret its consequences. What follows is not Clytemnestra telling instead of showing – a mere retelling of the completed action – but rather the Chours' consequent explanation of the scene before them and their announcement of things to come.

We can take Euripides' *Iphigenia at Aulis* as an example of a complex plot because it shares many of the same dramatis personae as *Agamemnon*. This tragedy is set in an earlier point during the Trojan War and it follows: Agamemnon's reluctant decision to sacrifice his daughter (Euripid, 1990, pp. 461–2) for the Hellenic victory; Achilles' attempt to save Iphigenia (Euripid, 1990, pp. 468–9, 472–3) despite the sabre-rattling Achaeans demanding immediate departure from Aulis to return to the battlefield (Euripid, 1990, p. 473); and Iphigenia's final voluntary decision to be sacrificed before her salvation (Euripid, 1990, pp. 473–6).

At the beginning of the tragedy, Agamemnon is trying to stop the sacrifice of his daughter. However, his message is intercepted by Menelaus. The reversal in the first two acts lies in Agamemnon's realization that if he does not sacrifice Iphigenia and lead the army back to the battlefield, Odysseus will arrive with his army and the number of casualties will increase dramatically (Euripid, 1990, pp. 461–2). Recognition in the fourth act occurs when Achilles and Clytemnestra are confronted with Agamemnon's plan and deception (Euripid, 1990, pp. 466–7). The tragedy ends with a final reversal, the most incredible of them all, when Iphigenia is spared, and a doe is sacrificed in her stead (Euripid, 1990, p. 476). Yet despite all of these changes and reversals, the plot retains unity – it focuses solely on the sacrifice of Iphigenia and the return of the Achaean army to the battlefield.

Whether the tragic plot is simple or complex, its task is the same – to achieve, "through pity and fear [the purification of] such emotions" (Aristotle, 2004, p. 244) – and for that it needs dramatis personae who, through action, drive the plot forward. Let us now examine this highlighted part of Aristotle's definition.

2.2. The evocation of fear and pity

Aristotle asserts that a tragedy imitates "events that evoke fear and pity. These effects occur above all when things come about contrary to expectation" (Aristotle, 2004, p. 249). In *The Rhetoric*, Aristotle defines fear "as a pain or disturbance due to a mental picture of some destructive or painful evil in the future." (Aristotle, 2012, p. 93) This alignes precisely with Achilles' words in *Iphigenia at Aulis* where he connects this emotion with the talk of the future:

> Your word spreads its wings into the future, hides a kind of fear. (Euripid, 1990, p. 466)

Aristotle points out that we cannot fear something that lies in the far future (Aristotle, 2012, p. 93) and that is, hence, uncertain. However, if we look at fear as a tragic emotion, we cannot claim that we are afraid of what will happen to us as we are watching the play or after we leave the theatre. Because of this, we must further enquire into this emotion.

> We also fear those who are to be feared by stronger people than ourselves: if they can hurt those stronger people, still more can they hurt us; and, for the same reason, we fear those whom those stronger people are actually afraid of. Also those who have destroyed people stronger than we are. (Aristotle, 2012, p. 94)

This quote, then, justifies the feeling of fear caused by the tragedy, "[s]ince tragedy is a representation of people who are better than we are" (Aristot. *Poet*. 1454b9–10).

Deretić emphasizes that Aristotle's interpretation of emotions in general, and of fear in particular, is, ultimately, cognitivist. Fear is not merely being afraid, "but it includes the belief that one is in danger and the estimation of a kind of danger which threatens one." (Deretić, 2019, p. 36) Nussbaum too, understands Aristotle's conception of pity as "based on thought and evaluation." (Nussbaum, 1996, p. 8) Aristotle himself asserts "that fear sets us thinking what can be done" (Aristotle, 2012, p. 95). This leaves the door open for fictional fear as well. Deretić indicates that tragic fear is what we experience while watching a tragedy, even though there is no real threat to us. By reading Greek tragedies or going to the theater, we agree to be aware that it is the tragic action that evokes feelings in us. Furthermore, we do not make a mistake when we assess the plot as fearsome. We can say that our fictional fear doesn't make mistakes (Deretić, 2019, p. 37).

Aristotle connects fear and pity and claims that, "[s]peaking generally, anything causes us to feel fear that when it happens to, or threatens, others cause us to feel pity." (Aristotle, 2012, p. 95) According to his words, we pity those similar to us, and fear for ourselves and those dear to us, because it is "the general principle that what we fear for ourselves excites our pity when it happens to others." (Aristotle, 2012, p. 105)

Another similarity between these two emotions is their temporal dimension. Namely, while the terrible deed is set in the near future which we anticipate, to feel pity, according to Aristotle, misfortune should occur either in the extremely near future, or it should have just happened (Aristotle, 2012, p. 105).

> Pity may be defined as a feeling of pain caused by the sight of some evil, destructive or painful, which befalls one who does not deserve it, and which we might expect to befall ourselves or some friend of ours, and moreover to befall us soon. In order to feel pity, we must obviously be capable of supposing that some evil may happen to us or some friend of ours . . . It is therefore not felt by those completely ruined, who suppose that no further evil can befall them, since the worst has befallen them already; nor by those who imagine themselves immensely fortunate (Aristotle, 2012, p. 103).

Importantly, "[p]ity occurs only when one is capable of putting oneself in the place of others." (Deretić, 2019, p. 38)

Another prerequisite for this feeling is to deem at least some people – those who were affected by that evil – good and undeserving of that evil (Aristotle, 2012, pp. 103–4). Deretić, hence, points out a normative dimension of this emotion. Namely, the judgement of the undeserved sufferings is expected from those who feel pity. This emotion requires us, on the one hand, to recognize ourselves as vulnerable, and, on the other hand, to have a certain distance from the object (person) of our pity. If it so happens that the person suffering is too close to us, we will not feel pity, but rather fear, even terror (Deretić, 2019, pp. 37–8). On the other hand, Nussbaum also speaks of the normativity of pity, but in a slightly different vein. Her stance is that

> all [pity] is "rational" in the descriptive sense in which that term is frequently used – that is, not merely impulsive, but involving thought or belief. Nevertheless, not all [pity] is rational in the normative sense, that is, based upon beliefs that are true and wellgrounded. (Nussbaum, 1996, pp. 30–1)

And when Rorty says that "[p]ity involves both distance and proximity" (Rorty, 1992, p. 12), the distance indicates that the people for whom we feel this emotion are not our closest friends and family, while proximity indicates our similarity to the sufferer, be that similarity in, as noted by Aristotle, age, social class or the context in which we live in (Aristotle, 2012, p. 105).

> While retaining awareness of her separateness, however, the pitier at the same time acknowledges that she has possibilities and vulnerabilities similar to those of the sufferer. She makes sense of the suffering by recognizing that she might herself encounter such a reversal; she estimates its meaning in part by thinking about what it would mean to encounter that herself, and she sees herself, in the process, as one to whom such things might in fact happen. That is why pity is so closely linked to fear, both in the poetic tradition and in philosophical accounts (Nussbaum, 1996, p. 35).

In short, there are, as indicated by Nussbaum, three necessary beliefs that form the foundation of pity, without which we would have no ground to feel this emotion:

> (1) the belief that the suffering is serious rather than trivial; (2) the belief that the suffering was not caused primarily by the person's own culpable actions; and (3) the belief that the pitier's own possibilities are similar to those of the sufferer. (Nussbaum, 1996, p. 31)

But it also seems that there is a difference between fear and pity. It is as if we who feel pity should ourselves possess certain moral qualities. Benn points out the selfelessness of pity (Benn, 1914, p. 87), but also indicates "the machinery of character-drawing [in the Greek tragedy] by which pity is converted into selfish terror [i.e. fear]" (Benn, 1914, p. 88), which is something Aristotle himself also notices (Benn, 1914, p. 88).

3. The turn in the tragic hero's fortune

The evocation of fear and pity is accomplished by following the plot of the tragedy played out by the dramatis personae. However, there are certain conditions that should be met in order for this effect to occur. According to Aristotle, the decline of decent men from fortune to misfortune doesn't evoke tragic emotions in us, but rather disgust. The change of wicked men's misfortune into fortune violates the assignment of the tragedy and, not only does it not evoke tragic emotions in us, but it doesn't evoke the feeling of philanthrophy either. And even when those utterly wicked experience the change of fortune into misfortune, we do not feel fear and pity (Aristot. *Poet*. 1452b34–36), "for [pity] is [felt] for a person undeserving of his misfortune, and [fear] for a person like [ourselves]." (Aristot. *Poet*. 1453a4–5) Still, we can note that as a just scenario.

Tragic emotions are evoked in us by

> the person intermediate between these. Such a person is one who neither is superior [to us] in virtue and justice, nor undergoes a change to misfortune because of vice and wickedness, but because of some error, and who is one of those people with a great reputation and good fortune, e.g. Oedipus, Thyestes (Aristot. *Poet*. 1453a7–12).

And, truly, King Oedipus is similar to us, though not in his class or his wisdom, but in his humanity and fallibility. Already in the second act, he is faced with the truth (Sofoklo, 1990, p. 134) which is painful and awful, like looking directly into the Sun immediately upon exiting the cave (Plat. *Rep*. 7.515e-516a). Because of that, the king, unsurprisingly, first points the finger at Creon for causing him to send for the prophet Tiresias, which to Oedipus seems like a conspiracy (Sofoklo, 1990, pp. 135–8). In the end, after he himself understands what truly happened (Sofoklo, 1990, pp. 139–140, 143–6) and after the tragedy's crucial recognition and reversal occur, Oedipus blinds himself (Sofoklo, 1990, p. 147).

3.1. Virtue

Just as there is only this intermediate man, the one similar to us, who can evoke fear and pity in the audience, moral virtue too is, according to Aristotle, that which lies between two vices – the one of excess and the one of lack. Virtue is the product of finding this right measure in feeling and in action (Aristot. *NE* 1109a20–21). Thus, it is feelings that form virtue, and so, we can seek the right measure in their expression. Where a person feels neither excessive boldness nor panic, one can feel fear and, therefore, pity (depending on the object of feeling) (Aristotle, 2012, p.

104). Rationally justified fear can be formed in that right measure of the relation between different feelings, as we have already seen that emotions have a cognitive dimension in Aristotle's opinion. With that healthy emotion as a foundation, Aristotle stresses, we can cultivate it and form a habit to act in specific ways in fearsome situations, thus forming virtue. Habit formed in this case will therefore be the built virtue of courage (Aristot. *NE* 1103a14-b25).

Still, let us take a step back in order to try to explain the true relationship between emotions, dispositions and virtue. When analyzing virtue, Aristotle asserts that there are three different states of our soul: emotions, capacities and dispositions (Aristot. *NE* 1105b20–3). Before proceeding, we must make certain terminological clarifications. Namely, in this paper, the word 'disposition' will be used for what Aristotle calls capacity, and the word 'habit' for what Aristotle calls disposition. With that in mind, we can say that we have dispositions to feel certain emotions. If our dispositions towards emotions are manifested time and time again, we can cultivate a habit, either one of virtue or one of vice. Dispositions, then, are the bridges connecting emotions with habits (Aristot. *NE* 1105b20–1106a). Perhaps we can imagine an emotion as a person walking down the road (disposition), in order to arrive at the final destination (habit). And if the building to which the person has arrived is ruined, the habit cultivated is that of vice. But if one stands before a beautiful mansion, we can say that it is virtue that is cultivated.

Dispositions, thus understood, belong to *hexis*, as noted by Slaby and Wünscher. But they also notice that "the term *hexis* is not limited to the human sphere alone" (Slaby, Wünscher, 2014, pp. 217–8). With that in mind, we can say that tragedy too can have dispositions, and that it can be good or bad depending on how it manifests those dispositions. Slaby and Wünscher also see *hexis* as a bridge that

> does *not* involve moral *reasoning* or theoretical reflection. By our *hexeis* we are already well– or ill-disposed *to act* in the light of strong emotions or passions. By virtue of our courage, for example, we are well-disposed to strike when facing a threat; but this does not mean we will not *feel* fear; it means that we will *turn* our fear around into brave action, and in this action we can, so to speak, *own* our fear. This should not be misconstrued as the commonplace ability to act in various ways *out of* our emotions. What Aristotle aims for is the ability to mould the emotion itself in accordance with our capacities and the situation at hand. This 'moulding' or shaping the emotion is carried out by acting appropriately (Slaby, Wünscher, 2014, p. 218).

From all of this, we can see that the bond between feelings and actions is quite strong. The dramatis personae in action are propelled to do something by certain emotions, just as is the case with us. Still, that is not the only way in which virtue can be manifested in a tragedy. Namely, the "principal virtue and justification" of a complex plot, Else notices, "is the enhancement of the *emotional content* of tragedy" (Else, 1957, p. 328). We can thus conclude that the goal and purpose of tragedies are to evoke tragic emotions, but also that we will judge those best – the most virtuous – tragedies most highly which have successfully accomplished that goal.

We see virtue and mistake (or its lack) at work in Sophocles' *Antigone*. Aristotle points out how grand Antigone's act is when she decides to bury her brother and follow the natural (divine) law, instead of the state (man-made) law (Aristotle, 2012, pp. 63, 69). Her deed, although connected to sin, is actually in stark contrast to it:

> So be as you decide to be – but I
> Will bury him. For me it's noble to do
> This thing, then die. With loving ties to him,
> I'll lie with him who is tied by love to me,
> I will commit a holy crime, for I
> Must please those down below for a longer time
> Then those up here, since there I'll lie forever. (Sophocles, 2007, p. 56)

Antigone as a dramatis persona is an exception. Her decline from fortune to misfortune doesn't occur due to a mistake, but, as indicated by Benn, her fall is in an act of heroic virtue. "The act of Creon in forbidding the burial of Polyneices is not an error; it is a great crime against the divine law and for that reason is justly visited with the divine vengeance." (Benn, 1914, p. 88)

It should be, however, pointed out here that, as Konstan observes, the opposing sides in tragedy should not be so diametrically "different in respect to virtue; it should follow, although Aristotle does not say so, that an audience can symphatize with both sides, provided that neither acts out of ignorance or sheer malevolence." (Konstan, 1999, p. 3) The reason behind the exclusion of perfectly moral and evil people is to ensure that the question of just and adequate punishment can even be asked (Benn, 1914, p. 88).

3.2. Hamartia

The aforementioned makes us return to the question of the change "*from* good fortune *to* bad fortune . . . due . . . to a serious error" (Aristotle, 2004, p. 251), i.e. of *hamartia*. The audience will relate to those dramatis personae, since they themselves could make such errors (Benn, 1914, p. 88). Konstan indicates that we do not take a vicious man as a tragic hero because the change in his fortune from good to bad, if justified, wouldn't be able to evoke tragic emotions in us, and because the audience doesn't relate to such a man (Konstan, 2006, pp. 214–5, 218). "[T]he condition that pity is a response not to any misfortune, or to misfortune that one has brought upon oneself, but only to undeserved hardship, provides the link between pity and justice." (Konstan, 2006, p. 204) It is necessary to notice here that we should, thus, take good people as tragic heroes. Else points out that the proper way to interpret this would be to say that "'[g]ood' does not mean 'perfect, faultless', but on the other hand Aristotle does not allow for any serious moral fault. His hero is to be a good man, though subject to error like all of us" (Else, 1957, p. 457). Despite this, his "tragic error is "big" (*μεγάλην*, 13. 53a16) in its dimensions and consequences *for the character of the play*. It leads to death, wounding, blinding, or the like: i.e., the *pathos*." (Else, 1957, p. 189)

Authors notice that the tragic concepts of recognition and fatal flaw are closely related. However, while the change in fortune is necessary for a tragedy, recognition and reversal are not (Daniels, Scully, 1992, p. 208). Yet, as Else points out, "[*h*]*amartia* must be a functional element in a complex plot" (Else, 1957, p. 379), and we have already said that the tragic plot is complex when it contains recognition and/or reversal.

> [T]ragic recognition, or the best tragic recognition, is a discovery of the identity of a 'dear' person, a blood-relative; it follows that the precedent *hamartia* would denote particularly a mistake or error or ignorance as to the identity of that person. (Else, 1957, p. 379)

Let us now, using examples, examine the tragic error and its implications.

Else highlights the following. Namely, if we wish to check whether the deed done in ignorance happened unintentionally or not, we can rely on the later presence or lack of remorse. If there is remorse after the deed done in ignorance, we have a clear, even necessary indicator that the act was unintentional. Else even goes a step further in thinking that we cannot claim that, if there is no remorse after the deed was done in ignorance, the situation wouldn't be different if information were known. Thus, this kind of act isn't caused by ignorance. Furthermore, the true acts done in ignorance are those caused, not by the lack of knowledge of general principles, but by not knowing the particulars – those plots evoke tragic emotions in us (Else, 1957, p. 380). *Hamartia* is a significant error that appears when the dramatis persona who is doing the act lacks the knowledge of certain details, ideally those concerning the identity of someone close to it (Else, 1957, p. 383).

> As a component or cause of the complex plot, such a *hamartia* is inherently fitted to arouse our pity – and our 'fear', that is, our horror that a man should have killed or be about to kill a 'dear one'. The discovery is then the counterpart and reverse of the mistake. Here the emotional charge which is inherent in the mistake (not in the ignorance *per se*, but in the horrible deed to which it stands in causal relation) finds its discharge. The *hamartia* represents the reservoir of emotional potential, the recognition is the lightningflash through which it passes of. (Else, 1957, p. 383)

However, let us examine the case of Euripides' *Medea*. Already in the opening verses (Euripid, 1990, p. 438), we are invited to pity Medea. Yet the case of *hamartia* in this tragedy is extremely atypical.

> If the agent is to blame, it is not for what he did, but because he did not resist with enough vigour. What the agent found himself doing had a reason – the passion or impulse that overcame his better judgement – but the reason was not *his*. From the agent's point of view, what he did was the effect of a cause that came from outside, as if another person had moved him. (Davidson, 1982, p. 295)

According to this interpretation, *hamartia* is not ascribed to Medea, but to Jason. Schauroth also noticed this. Medea seems like a simple instrument in the plot already set in motion by Jason's decision to remarry. Schauroth notices the similarity between this tragedy and Aeschylus' *Agamemnon*. Namely, in both cases, the plot is propelled by the husbands' acts, while the wives serve as the means of vengeance. In both cases, it is the husbands who, unbeknownst to them, made the mistake. They are the ones whose fortunes change and who suffer because of their mistake (Schauroth, 1932, pp. 366–7). However, both Clytemnestra and Medea lost a great deal due to their own deeds, and it is hard, to say the least, not to consider them charismatic.

Still, what is characteristic of typical tragedies is their ability to evoke tragic emotions in us. Else, rightly poses the question "under what circumstances can I bring myself to pity the murderer or would-be murderer?" (Else, 1957, p. 437) He answers this question by saying:

> I can pity him if I judge that *he did not intend the parricide, matricide, or whatever, as such*: in other words, if it is established to my satisfaction that he performed or intended the fearful act . . . because of some error. (Else, 1957, p. 437)

Similarly, Nussbaum states:

> Putting seriousness and fault together, we see that pity requires the belief that there are serious bad things that may happen to people through no fault of their own, or beyond their fault. In pitying another, the pitier accepts a certain picture of the world, according to which the valuable things are not always safely under a person's control, but can be damaged by fortune. (Nussbaum, 1996, p. 33)

4. Tragic emotions

After examining the reversal in the tragic hero's fortune, let us return to the definition of tragedy.

> Tragedy is an imitation of an action that is admirable, complete and possesses magnitude; in language made pleasurable, each of its species separated in different parts; performed by actors, not through narration; effecting through pity and fear the purification of such emotions. (Aristotle, 2004, p. 244)

The interests of this paper are the tragedy's aspects of content, and not these formal ones (i.e. completeness, proper magnitude). For that reason, until now, we've dealt with the question of the dramatis personae in action, and now we are moving to the investigation of the tragedy's main task – achieving catharsis. In order to do so, we should once more deal with fear and pity.

The significance of evoking pity and fear for a tragedy is, as can be seen from the fact that it is included in the definition, enormous. These emotions are "the defining characteristics of the tragic genre." (Deretić, 2019, p. 39) Following Rorty and Halliwell, Deretić notices that the suffering and change in the tragic hero's fortune – reversal and recognition – are that which evokes pity and fear. However, in order for these emotions to be evoked, it is necessary that the structure of the plot itself is plausible (Deretić, 2019, p. 39). Not only that, but, as Ben-Ze'ev notes, the more plausible and believable the situation is, the more intense the emotions are (Ben-Ze'ev, 2001, p. 126).

Until now, fear was viewed as a selfish emotion – we fear that fearsome thing which can befall us or our closest ones in the near future. However, when put in the context of a tragedy, this emotion "occurs when the audience experiences the disturbing and painful feeling concerning the calamities that might happen to the protagonists as they are dramatized in a tragedy." (Deretić, 2019, p. 40) It is as if the borders of proximity imposed on us by fear and pity are suddenly moved, now including people who, not only are not our close family and friends, but may be thousands of years away from us. Deretić highlights that this became possible because we relate to the tragic heroes to some extent. It is precisely because we recognize this similarity and understand that the fearsome things could befall not only tragic heroes, but us as well, that we feel fear. "Tragic fear is grounded in the belief that we understand and sympathize with a tragic hero's calamities, as if they were our own." (Deretić, 2019, p. 40) Deretić points out that this fear, not only is no longer selfish, but it's as if it demanded to be transformed into some sort of altruistic fear, one that would, on the basis of sympathy with others, encompass them as well. Apart from that, the author refers to Ferrari (Deretić, 2019, pp. 40–1), who claims that it is necessary for the audience to start to "care enough about the hero to fear for him, that is to say fear on his behalf." (Ferrari, 1999, p. 196) This goes to show that the audience has formed a certain opinion about and emotional connection to the dramatis persona, the theatregoers have connected to it, and they relate to it to such an extent that the audience starts to feel what the dramatis persona itself ought to feel.

On the other hand, Benn argues that, despite the fact that our feelings towards tragic heroes should be disinterested, like sympathy (Benn, 1914, p. 86), and while their fatal mistakes should be such that any audience member could have made (Benn, 1914, p. 88), tragedy's purpose is not to suggest that such acts are commonplace. He contends that Aristotle's examples demonstrate how unlikely it is that the average theatergoer would repeat Oedipus' actions. Were the viewer in place of the tragic heroes, it is unlikely he would act like them. For that reason, Benn believes that this fear does not reflect real possibilities. By watching a play, pity elevates to terror, and yet precisely because the tragedy depicts extraordinary events, this fear returns to the level of sympathy and thus neutralizes itself (Benn, 1914, p. 89). On the other hand, Deretić notices that our problems appear trivial compared to those of tragic heroes (Deretić, 2019, p. 41).

> It might be said that watching tragedies refines and cultivates our emotion of pity by making us more sensitive and vulnerable to tragic human calamities. Due to the fact that it is cognitive in character, pity deepens our understanding of others and our relations to them, in this way contributing to our self-understanding. (Deretić, 2019, p. 41)

However, despite the differences between ordinary and tragic pity regarding who we can feel this emotion for, there are numerous similarities between them, too (Deretić, 2019, p. 41). Frijda tries to base that on our concepts of "reality" and "unreality". We can slightly modify his examples to better fit our topic. For instance, if we were to witness an imaginary eagle approaching an imaginary Prometheus, that situation would have a completely different impact on us than one in which a fake eagle attacks real Prometheus. "The unrealities are of a different order. The first is seen as real in an imaginary world, and the second as unreal in a real one." (Frijda, 1989, p. 1546) If we, as the tragedy's audience, leave our 'reality' behind and accept our place in the game of make-believe, or, as Frijda says, willingly suspend our disbelief, we can feel aesthetic emotions (in this case, tragic emotions) that do not differ from our usual emotions. They are just transponded to the plane of the tragedy.

> In both cases, the emotions stem from the same source: one's power to identify with someone else and to see what befalls her or him from her or his angle, while at the same time knowing that it is not oneself who is concerned. In both cases, too, the distinction between the other individual and oneself may get blurred. (Frijda, 1989, p. 1546)

Nussbaum takes the example from Sophocles' *Philoctetes* where The Chorus imagines and pities the man they don't know, just as we, the audience, are called to do upon entering the theatre (Nussbaum, 1996, p. 27). We have seen that distance is necessary for pity, even in the real world, while fear is characterized by proximity. For that reason, it is not hard to feel pity for dramatis personae, despite the ontological gap that exists between them and us. Deretić points out that we feel both kinds of pity for those suffering due to extremely painful and undeserved adversity. In both cases, we pity the consequence of the mistake from which there is "no going back" – a mistake that cannot be fixed (Deretić, 2019, pp. 41–2).

Aristotle's system is characterized by the understanding that

> [t]he objects of both [tragic] emotions are serious harms . . . The emotions in question entail evaluation, though pity, that has a richer cognitive structure, presupposes the moral judgement of what are "undeserved sufferings", grounded in a basic sense of justice. (Deretić, 2019, p. 42)

And although the mistake made by the tragedy's protagonist cannot be fixed, there is still a possibility that we ourselves would have acted differently in that situation. Hence, watching or reading a tragedy can show us, Ben-Ze'ev notes, "alternatives we didn't think existed. Encountering such new alternatives excites us." (Ben-Ze'ev, 2001, p. 129)

4.1. Dispositions to tragic emotions

Everything discussed thus far can be set in a dispositional framework which will, in turn, guide us towards a better understanding of the task of tragedy and the conditions under which it is successfully carried out. Based on the previous considerations, tragic emotions can now be understood as manifestations of dispositions to empathize with characters facing great suffering and misfortune. Given that, we can rephrase the definition of tragedy as follows:

> Tragedy is an imination of an action that is admirable, complete and possesses magnitude; in language made pleasurable, each of its species separated in different parts; performed by actors, not through narration. It has a disposition to, effecting through pity and fear, purify of such emotions.

A definition phrased in this way would indicate more precisely that the task of tragedy is to evoke tragic emotions in order to achieve catharsis. Hence, while not every tragedy has to accomplish this effect, we consider a successful tragedy to be the one that does. Thus, we can understand catharsis as the final result of the manifestation of dispositions to tragic emotions.

The tragedies' dispositions to tragic emotions could be explained even better if we take a look at the following. Namely, Barwell begins his inquiry into the question of how art expresses emotion by examining sentence schemata that address the relationship between an artwork and an emotion (Barwell, 1986, p. 175). So, for instance, we can take certain sentences to mean that – using tragic emotions as examples – a tragedy has a tendency [disposition] to arouse pity in an attentive audience. Saying that a tragedy is scary would mean that it arouses fear in its audience. "Any artwork which is expressive of an emotion has a tendency [disposition] to evoke the emotion *in the imagination* of an attentive audience, but this evoking need not be emotionally loaded." (Barwell, 1986, p. 177) This would mean that, by calling a tragedy scary, we recognize that it has a disposition to arouse fear and that it can, in turn, evoke fear in its audience.

The disposition toward tragic emotions found in Greek tragedy is all the more understandable when we consider Ben-Ze'ev's following words:

> The more we know about a certain object, the more real (in both senses) it is typically perceived to be. An object perceived by sight, hearing, and smell is usually considered to be more real than that perceived by sight alone—other things being equal. More detailed knowledge about someone's life makes this person more real to us than someone we hardly know about. (Ben-Ze'ev, 2001, p. 130)

Namely, it is precisely because the average theatergoer of the time was well-versed in myths, and knew the context of the tragedies beforehand that the audience found tragedy real enough. Since it didn't have to "waste time" on the introduction of the dramatis personae or even the plot itself, Greek tragedy's focus is on manifesting dispositions to tragic emotions by focusing on pitiful and fearful elements.

4.2. Catharsis

The last part of the definition we will examine is catharsis, which is arrived at after tragedy accomplishes "the purification of [tragic] emotions" (Aristotle, 2004, p. 244) by evoking said emotions.

As stated by Deretić, the previously established point of view according to which catharsis literally means the physiological and behavioral manifestation of tragic emotions has largely been abandoned today (Deretić, 2019, p. 48). This is completely understandable. The cases of violent reactions (e.g., unstoppable stream of tears, screaming) to artworks are extremely rare. For that reason, it seems unconvincing to interpret the task of tragedy as a way to get accumulated negative emotions "out of the system". Deretić indicates that the tragedy, according to this view, would funciton to enhance our emotions, and then release them, while its content, multidimensionality, and the questions it poses would be neglected. This last part is crucial, as it is precisely on this basis that representatives of the second group of exponents of catharsis maintain that tragedy's goal is the clarification of universal human behaviors through representing particular actions on stage. Interpretation of this kind demands mental involvement from the audience. Apart from that, this interpretation could justify the need to feel tragic emotions, although they are usually viewed as negative (Deretić, 2019, pp. 49–50).

> Their [defenders' of Intellectual Interpretation] explanation of tragic pleasures is in accordance with Aristotle's insistence that the pleasure specific to tragedy arises from pity and fear through *mimesis* (*Poet*. 1453b10–13). More importunately, by determining the tragic pleasures as the intellectual ones, this model of interpretation successfully explains how unpleasant feelings like pity and fear can be transformed into pleasant enjoyments. By watching a drama, spectators take pleasure in understanding "the pitiful and fearsome incidents" in a clearer and subtler manner. Primarily, they can infer what is universal in these particular tragic events (Golden 1962: 53–55). Additionally, they might find out how world and human relationships are even more complex and fragile then they previously thought. (Deretić, 2019, pp. 50–1)

This approach places chatarsis in the tragedy itself and interprets it as a property of tragedy (Deretić, 2019, p. 51). Such an understanding of the notion, as well as other interpretations that view chatarsis as a property of tragedy, aligns perfectly with the proposed dispositional framework for analyzing tragic emotions.

In this vein, Nehamas redefines tragedy as follows:

> Tragedy, then, is an imitation of action that is serious, complete, and of a certain magnitude, using language embellished in various ways in its differ-

> ent parts, through action and not through narration, and carrying such incidents to their appropriate resolution through a course of events that provoke pity and fear. (Nehamas, 1994, p. 280)

Nehamas asserts that the tragic plot itself is thus responsible for evoking tragic emotions. “But though pity and fear are also products of drama, they are primarily part of its content” (Nehamas, 1994, p. 280). Hence, it is the tragic plot itself that must be fear-inducing and pitiable. The audience could, then, recognize the manifestation of the disposition toward tragic emotions.

I should distance myself from this attitude, and mention that Deretić pointed out the contemporaneity of such interpretations of catharsis and, in general, of Aristotle’s theory of tragedy (Deretić, 2019, p. 52). This precisely constitutes an attempt to provide a single interpretation, and not a translation of Aristotle’s text.

5. Conclusion

This paper aimed to present an interpretation of Aristotle’s understanding of tragedy. The texts used for this purpose were the *Poetics*, the *The Art of Rhetoric*, and, to a lesser extent, the *The Nicomachean Ethics*, to provide the reader a broader context of his understanding of certain concepts.

In the first part of the paper, we dealt with a preliminary analysis of Aristotle’s definition of tragedy. After this definiton was introduced, we placed special emphasis on the questions of dramatis personae in action and the evocation of pity and fear, i.e. tragic emotions. The second part of the definition, which we initially paid attention to, considered the emotions of pity and fear and offered their initial interpretations.

The second part of the paper concerned itself with changes in tragic heroes’ fortunes. Certain preliminary remarks were reiterated in order to connect these changes to Aristotle’s understanding of virtue and to further elaborate on the clarification of *hamartia*, i.e. fatal mistake of the tragic hero that propels the plot of the tragedy. Within these two parts of the paper, representative examples were presented to provide additional clarification of universal ideas through particular examples.

In the third last part of the paper, we returned once more to the definition of tragedy, this time in order to investigate tragic emotions, i.e. tragic pity and tragic fear. Here, we examined these two tragic feelings, and presented their similarities and differences compared to the initial interpretations of these emotions. Additionally, we provided an interpretation of all preceding points by placing them within the dispositional conceptual framework. The idea was to show that the goal of tragedy is to evoke tragic emotions that, in turn, lead to catharsis. To accomplish this, we sought to explain tragic emotions as properties of tragedy. Lastly, we examined certain interpretations of catharsis and found that some can be compatible with the dispositional framework. There, chatarsis itself was viewed as one of the properties of tragedy for the sake of harmonizing it with the proposed dispositional framework.

This paper aimed solely to provide the reader with an interpretation of Aristotle's thought, not a translation. Fear and pity, *hamartia*, the reversal in the fortunes of the dramatis personae who act, and catharsis are all shown to be legitimate properties and elements of tragedy. The manifestation of tragedy's disposition toward tragic emotions determines it success and justifies evaluating it as such, as the examples support.

Bibliography

Aristotel (2015). *O pesničkoj umetnosti*. Beograd: Dereta.

Aristotle (2004). Extracts from *Poetics* by Aristotle. In S. Dutta (ed.), *Greek Tragedy* (pp. 239–255). London: Penguin Books.

——— (1987). *Poetics*. Indianapolis & Cambridge: Hackett Publishing Company.

——— (2012). *The Art of Rhetoric*. London & Dublin: Collins Classics.

——— (1996). *The Nicomachean Ethics*. Hertfordshire: Wordsworth Classics of World Literature.

Barwell, I. (1986). How Does Art Express Emotion? *The Journal of Aesthetics and Art Criticism*. 45(2), 175–181.

Benn, A. W. (1914). Aristotle's Theory of Tragic Emotion. *Mind*, 23(89), 84–90.

Ben-Ze'ev, A. (2001). *The Subtlety of Emotions*. Cambridge & London. The MIT Press.

Currie, G. (1993). Review. *The Journal of Philosophy*. 90(7), 367–370.

Daniels, C. B. and Scully, S. (1992). Pity, Fear, and Catharsis in Aristotle's Poetics. *Noûs*. 26(2), 204–217.

Davidson, D. (1982). Paradoxes of irrationality. In R. Wollheim and J. Hopkins (eds.), *Philosophical Essays on Freud* (pp. 289–305). Cambridge: Cambridge University Press.

Deretić, I. (2019). Why is Tragedy Good for the Soul? Aristotle on the Tragic Emotions and *katharsis*. In Lj. Radenović, D. Dimitrijević and Il Akkad (eds.), *Pathe: The Language and Philosophy of Emotions* (pp. 33–58). Belgrade: University Library.

Else, G. F. (1957). *Aristotle's Poetics: The Argument*. Cambridge, Massachusetts: Harvard University Press.

Eshil (1990). Agamemnon. U V. Topalović i B. Brkić (eds.), *Sabrane grčke tragedije* (pp. 59–80). Beograd: Vrhunci civilizacije.

Euripid (1990). Ifigenija u Aulidi. U V. Topalović i B. Brkić (eds.), *Sabrane grčke tragedije* (pp. 455–477). Beograd: Vrhunci civilizacije.

——— (1990). Medeja. U V. Topalović i B. Brkić (eds.), *Sabrane grčke tragedije* (pp. 437–454). Beograd: Vrhunci civilizacije.

Ferrari, G. R. F. (1999). Aristotle's Literary Aesthetics. *Phronesis*. 44(3), 181–198.

Frijda, N. H. (1989). Aesthetic Emotions and Reality. *American Psychologist*. 44(12), 1546–1547.

Konstan, D. (2006). *The Emotions of the Ancient Greeks: Studies in Aristotle and Classical Literature*. Toronto, Buffalo & London: University of Toronto Press.

——— (1999). The Tragic Emotions. *Comparative Drama*. 33(1), 1–21.

Nehamas, A. (1994). Pity and Fear in the *Rhetoric* and the *Poetics*. In D. J. Furley and A. Nehamas (eds.), *Aristotle's "Rhetoric": Philosophical Essays* (pp. 257–282). Princeton: Princeton University Press.

Nussbaum, M. (1996). Compassion: The Basic Social Emotion. *Social Philosophy and Policy*. 13(1), 27–58.

Platon (2017). *Država*. Beograd: Dereta.

Rorty, A. O. (1992). The Psychology of Aristotelian Tragedy. In A. O. Rorty (ed.), *Essays on Aristotle's* Poetics (pp. 1–22). Princeton: Princeton University Press.

Schauroth, E. G. (1932). Some Observations on Aristotle's View of Tragedy. *The Classical Journal*. 27(5), 352–368.

Slaby, J. and Wüschner, P. (2014). Emotion and Agency. In S. Roeser and K. Tod (eds.), *Emotion & Value* (pp. 212–228). Oxford & New York: Oxford University Press.

Sofoklo (1990). Kralj Edip. U V. Topalović i B. Brkić (eds.), *Sabrane grčke tragedije* (pp. 129–150). Beograd: Vrhunci civilizacije.

Sophocles (2007). *Antigone*. Oxford & New York: Oxford University Press.

7. DISPOSITIONS IN AESTHETICS AND ETHICS

https://doi.org/10.18485/baf.2025.2.ch14

Aleksandar Risteski

THE ANTHROPOLOGICAL IMPLICATIONS OF PEIRCE'S AESTHETICS*

Abstract: In this paper, I will focus on Peirce's views on aesthetics and some of its anthropological implications. Though Peirce never developed a systematic account of aesthetics, he was aware of its importance for ethics and logic (C.P. 8.225–8.256). The paper addresses Peirce's notion in C.P. 5.314–315, that a "man is the thought", from the view of his aesthetics. My intention is to show that Peirce's views on the "qualities of feeling" and on aesthetic experience in general may have interesting philosophic-anthropological implications.

Keywords: aesthetics, Peirce, philosophical anthropology, qualities of feeling, values.

1. Peirce's attitude towards aesthetics

1.1. Introduction

Discussing the nature of our conduct, Aristotle makes the following remark in *The Nicomachean Ethics*:

> "Every art and every inquiry, and similarly every action and choice, is thought to aim at some good; and for this reason the good has rightly been declared to be that at which all things aim. But a certain difference is found among ends; some are activities, others are products apart from the activities that produce them. [...] If, then, there is some end of the things we do, which we desire for its own sake [...] clearly this must be the good and the chief good. Will not the knowledge of it, then, have a great influence on life? Shall we not, like archers who have a mark to aim at, be more likely to hit upon what is right? If so, we must try, in outline at least, to determine what it is, and of which of the sciences or capacities it is the object. It would seem to belong to the most authoritative art and that which is most truly the master art." (Aristotle, 2009, pp. 3–4; 1094a1–25).

* This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03–136/2025–03/200184).

Aristotle recognizes a hierarchy of the ends of purposeful activities. Given the hierarchy, he assumes the existence of a highest good, or one worth pursuing for its own sake. The quoted passage does not fully reflect the main topic of this paper, but I find some parts of it interesting and relevant.

Peirce also recognizes the teleology of our conduct (cf. C.P. 6.155–6, 434; 7.570). He discusses the relationship between the "logically good", "morally good" and "esthetically good", or the ends we can ascribe to the phenomena of our experience (C.P. 5.120–150; cf. 1.191ff). He also employs the concept of teleology when discussing the nature of *personality* (cf. C.P. 6.155). However, Peirce's concept of teleology differs from the Aristotelian concept of natural, innate teleology (cf. Aristotle, 2018, pp. 21ff; 193a30ff; Shields, 2023). It also differs from Kant's views, because it is not reduced to how we tend to observe phenomena (cf. Kant, 2007, p. 21ff).[1]

Peirce's view on the teleology of human conduct is related to his understanding of *aesthetic experience*. How we form ideals we pursue is related to it (cf. C.P. 6.155–6, 434; 7.570). That is why ethics must rely on aesthetics, to determine the nature of the *summum bonum* (C.P. 1.191). This does not mean that the task of aesthetics is to definitively determine and disclose positive information on the content of the highest end.[2] What Peirce probably meant is that it should reflect on our tendency, due to the teleological structure of our actions, to posit something *as* the *summum bonum*. Irrespective of the content of an ideal, we act *as if* something is the highest good.[3] In that regard, we may assume, ethics should see aesthetics as an ally[4] in determining the correct mechanisms of our conduct.

The task of this paper is to assess the anthropological implications of Peirce's aesthetics. What are the implications of Peirce's understanding of

1 More on Aristotle's and Kant's understanding of teleology, see (Petrović, 2016). Peirce uses the expression *developmental teleology*, contrasting it to the understanding of teleology as a mere realization of a predetermined purpose. Cf. C.P. 6.156.

2 Such content is hardly known positively. Cf. C.P. 5.433: "[T]he pragmaticist does not make the *summum bonum* to consist in action, but makes it to consist in that process of evolution whereby the existent comes more and more to embody those generals which were just now said to be *destined*, which is what we strive to express in calling them *reasonable*".

3 This does not mean that Peirce's views on *summum bonum* are relativistic. This observation refers to the fact that it is not uncommon for people to act upon some belief without having a clear understanding of it. People form a belief and act *as if* something is true, even though that need not be the case.

4 More specifically, in C.P. 1.191, Peirce says: "Ethics, or the science of right and wrong, must appeal to Esthetics for aid in determining the *summum bonum*". It seems that "esthetics" alone does not *decide* an ideal, the highest good, since there are only aesthetic goods, and no aesthetic "grades of excellence" (C.P. 5.132). There is no aesthetic "goodness" and "badness" (ibid).

aesthetic experience, for his views on personality? The presented task may invoke some confusion. Namely, it is questionable whether there is such a thing as "Peirce's aesthetics", let alone its "anthropological implications". I believe, however, that it is possible to discuss the topic and that we can even compare Peirce's views with, for example, those of Plato, regarding the question of the importance of aesthetic education. Before that, however, we have to pay attention to some other important moments.

I intend to address Peirce's attitude towards aesthetics. The other point I would like to address is Peirce's view on the *beautiful*. How does Peirce understand the beauty, and how does he relate it to our *acting* and *thinking*? Furthermore, how can we understand his view of beauty as a *quality*? Thematization of these moments may provide us with an interesting background to contrast some of Peirce's aesthetic and anthropological notions.

For example, the view that *man is the thought* (cf. C.P. 5.314–315) presents itself in a slightly different light, given the background. The view must also include that which constitutes the thought or that which precedes it.[5] This relationship has been illustrated by Peirce's theory of categories and various notions on aesthetic experience.

1.2. Peirce's aesthetics

Peirce never developed a systematic account on aesthetics (*esthetics*). In his writings, we notice some occasional reflections, or what Herman Parret recognizes as "fragments on aesthetic experience" (Parret, 1994, p. 179ff). However, these fragments provide us with an interesting interpretative framework that can help us understand Peirce's views.

Peirce studied aesthetics in his youth, and then only after a long period of neglecting it, later in his life, recognized its importance (Atkinson, 1982, pp. 1–3; Potter, 1997, p. 52ff). He confesses that he was not well acquainted with it (C.P. 1.191). He was, in his own words, "lamentably ignorant of it" (C.P. 2.120), since he had "terribly neglected it" (C.P. 2.199). As a logician, he was more interested in other topics (C.P. 2.197).

These statements, however, do not fully reflect Peirce's views. His early encounter with philosophy is marked by his studies of aesthetics[6] (cf. C.P. 2.197; Atkinson, 1982, pp. 1–3; Parret, 1994, p. 179; Potter, 1997, p. 52ff). Also, he was critical of the general understanding of it, suggesting a different approach (cf. C.P. 2.199ff). Concerning his views, and the subject

5 Similarly, Damir Smiljanić tackles the topic of the relationship between the pathic and the cognitive aspects of ourselves. See (Smiljanić, 2020).

6 As early as in 1855, Peirce read Schiller's *Aesthetic Letters*, and started studying Kant. See (Peirce, 1982, pp. 1–3).

of his criticism, we might say that he was, in a way, addressing philosophical topics as an “aestheticist”.

This does not mean that he observed everything through the lens of aesthetics, especially not in a traditional sense, which he was critical of. His aesthetics is firmly related to his metaphysics and the theory of categories. Its subject ought to be Firstness and the qualities of feeling (cf. C.P. 8.255–256). Peirce was, however, mindful of the importance of this category and its different instances.

He also shows sensitivity for the arts and aesthetic qualities related to them. In some cases, Peirce was approaching philosophical problems not only as an aestheticist, but an artist, or at least as someone capable of grasping what an artistic appreciation, creativity, spontaneity and inspiration may be.[7] Aside of being interested in the question of “poetic mood”, he was, in a way, *employing* it. Let us look at an example that Peirce provides us with. In C.P. 5.44, he says:

> Go out under the blue dome of heaven and look at what is present as it appears to the artist’s eye. The poetic mood approaches the state in which the present appears as it is present. Is poetry so abstract and colorless?

In Peirce’s view, the quoted passage addresses some of the inadequacies of Hegelian metaphysics, foremost the idea of *immediacy* as *the pure being* (cf. Hegel, 1977, p. 58ff). It is important to notice that the “artist’s eye” is not merely a metaphor. Peirce views the ability to observe and *experience* things as an artist as an indispensable element of philosophical thinking, or, more specifically, an integral part of phenomenological methodology. A passage (C.P. 5.42) slightly before the above-quoted, sheds some additional light in this regard:

> When the ground is covered by snow on which the sun shines brightly except where shadows fall, if you ask any ordinary man what its color appears to be, he will tell you white, pure white, whiter in the sunlight, a little greyish in the shadow. But that is not what is before his eyes that he is describing; it is his theory of what *ought* to be seen. The artist will tell him that the shadows are not grey but a dull blue and that the snow in the sunshine is of a rich yellow. That artist’s observational power is what is most wanted in the study of phenomenology.

---

7 Cf. C.P. 5.112: “I have gone through a systematic course of training in recognizing my feelings. I have worked with intensity for so many hours a day every day for long years to train myself to this; and it is a training which I would recommend to all of you. The artist has such a training; but most of his effort goes to reproducing in one form or another what he sees or hears, which is in every art a very complicated trade; while I have striven simply to see what it is that I see”.

To understand this remark, we have to bear in mind Peirce's understanding of sense perception. Namely, sense perception is *abductive*, that is, it involves cognition (cf. C.P. 7.630; Anđelković, 2006; Trajkovski, 2015, p. 146ff). It is why any ordinary person is inclined to report what *should be* seen, but only a trained artist tends to step back and acknowledge *what is present as it is present*.

The appraisal of the observational powers of an artist does not imply that artists are natural-born philosophers, or that art is in any way superior to philosophy, like German romanticists claimed (cf. Frank, 2008, pp. 19–20, 53ff; Gorodeisky, 2016). The quoted passage tells us not that artists are generally more capable of addressing philosophical problems, but that a philosopher must not be oblivious of the element of reality that the "poetic mood" is directly attached to: the immediate, the present presentness, the present as it is present, or the First (cf. C.P. 1.356–1.357, 547; 5.44).

The quoted passages should not be interpreted as claiming that artists are closer to apprehending the "essence of things", or having some advantage in their "conceptual determination", compared to any other man. Sense perception supposes a semiotic tridimensional relationship, the mediation of signs. The underlying mechanisms of perceiving things by an artist, and anyone else, are just the same. Peirce probably gives an example with an artist to point out that there are different *habits* in observing things. An artist, we may assume from the given examples, is aware that there are elements usually overlooked, that are nonetheless present. They are overlooked *because* we, so to speak, predefine what it is that we (should) observe.

The implications for phaneroscopy are that it demands a similar kind of attention ("observational power"), or noticing of all the elements that can be related to an observed phenomenon, irrespective of its ontological status, the elements that usually can be overlooked because we *assume how we should* observe the phenomenon.[8] Phenomenological observation, and the "artist's observational power", cannot be thus identified with the Cartesian idea of "immediacy", or "immediate knowledge".

Peirce's observations can help us define a conceptual framework for understanding his possible attitude towards aesthetics. When we pay closer attention, we notice at least two opposing approaches. On the one hand, Peirce says that he is not familiar with aesthetics, but on the other hand, he points out what it *ought to* be.

Why has Peirce developed a sentiment towards something he is not being completely familiar with? We also notice that Peirce "takes the word

8 For example, to differentiate phenomena as those that belong to our imagination, or those that belong to the outside world, between mental or corporeal, etc. Phaneroscopy intentionally distances itself from such differentiations.

of others that there is such a science" (C.P. 2.120). When discussing aesthetic matters, he questions the very existence of aesthetics.

If we try to unite different statements to clarify Peirce's possible stance, we face three mutually opposing points:

- Peirce is not well acquainted with aesthetics (cf. C.P. 2.120);
- Peirce is not satisfied with what aesthetics of his day presents itself to be (cf. C.P. 1.574; 2.199);
- Peirce only takes the "word of others" that such science exists (cf. C.P. 2.120).

These three points are in fundamental disagreement. They form, to use Rescher's expression, a sort of *aporetic cluster* (cf. Rescher, 1987, p. 288ff). However, an aporetic situation is avoided if the points are seen as dialectical correlates, or different instances of an understanding of aesthetics that has yet to be clarified. In this regard, the second statement, that expresses Peirce's sentiment of what aesthetics should be, seems most helpful.

Peirce was clear about not having developed a systematic account of aesthetics. From his point of view, however, it is questionable whether *anyone* had. What went under the name of aesthetics clearly did not meet his demands. Maybe Peirce was unacquainted with it because of its questionable existence?

We may assume that Peirce's attitude towards aesthetics stems from his philosophical convictions regarding its true nature. There is more to this topic, but to continue it would first be necessary to address the question of beauty. After that, we will get back to Peirce's views on aesthetics as the basic normative science. So, what is *beautiful* according to Peirce?

2. The quality of beauty

2.1. Beautiful, unbeautiful and καλός

In Peirce's view, the beautiful is the object of admiration. It is more similar to Kant's concept of *the sublime*, given that it is not confined within the concept of a *form* (cf. Kant, 2007, p. 75ff; Parret, 1994, pp. 182–183). The idea of an aesthetic quality is related to the phenomenological concept of Firstness. The First is to be understood as twofold: it is the first, like a quality of feeling which, as Peirce says, *colors our personality* (C.P. 6.155), and it is the last, like *an ideal* to pursue (cf. Parret, 1994, pp. 179–180). We may think of it as something that determines our teleological matrix of conduct. The element of firstness is something that, in a way, determines us. Our appreciations of things, our setting goals and achieving them, seem to expose some of that determination (cf. C.P. 1.574).

Peirce was critical towards the traditional views of beauty as form and harmony. In fact, he was critical of the term *beauty* as such, and its interpretations as the primary subject-matter of aesthetics (C.P. 2.199). Similarly to Plotinus, Peirce argued that harmony and symmetry are only conditions for beauty to appear, and not something identical to it (cf. Parret, 1994, p. 182ff).[9] Aesthetics, he argued, is *crippled* by the very understanding of it being the science of the beautiful (C.P. 2.199). The concept is only *a product of aesthetics*, whose primary subject-matter should be something else (ibid).

Describing something as beautiful is one of the ways we may admire something. But to admire something is to appreciate it for what it is, without "any ulterior reason" (C.P. 1.191). In other words, aesthetics studies that which is admirable and desirable in itself, as it is, without any other ground of justification.[10] And such conduct is not reduced to appreciating fine arts or natural beauty alone. Instead, Peirce notices a connection between admiring something and experiencing *aesthetic qualities*, or certain *qualities of feeling*. He understands *ideal* as a *habit of feeling* (C.P. 1.574).

His understanding of beauty, thus, was not formal, cognitivist or representational, but axiological (Parret, 1994, pp. 182–183; Smith, 1972, p. 22ff). The way we evaluate something is determined by our aesthetic experience (cf. C.P. 1.574). However, the adjective "axiological" does not imply *moral axiology*. There is no aesthetically "good" and "bad" (C.P. 5.127). Aesthetics recognizes all of its phenomena as *aesthetic values* or *qualities*, so even something appalling may appear "delicious" in its "perfection" (C.P. 5.127).

There are only *qualities*, and individual instances that embody them. So, in a way, we can talk about the aesthetic qualities of something being "ugly" (C.P. 2.199); being repelled by or attracted to something is related to how different qualities are embodied. To equate "good" and "beautiful", that is, "evil" and "ugly", is a moral evaluation of aesthetic qualities (C.P. 5.127).

9 On Plotinus' concept of beauty see (Plotinus, 2018, pp. 91–103; I.6.1.25ff; Miles, 1999, pp. 37–42; A. Smith, 2016, p. 15ff).

10 We may look at Peirce's account on aesthetic experience as sharing similar tendencies and motives with the philosophical debates on what constitutes and motivates moral action, characteristic for post-Kantian German philosophers, and motivated by Kant's ethical thought. More on these debates cf. (Milisavljević, 2006, p. 19ff).

It seems that Peirce radicalizes Kant's notion of the beautiful as that which is independent of its explanatory potentials and moral values (cf. Kant, 2007, p. 42ff). The quality of being beautiful is, as Kant would say, "independent of all interest" (Kant, 2007, p. 36ff). But noticing and experiencing beauty, and different embodiments of different aesthetic qualities is not solely a matter of recognizing their form, or "purposiveness without purpose" (cf. Kant, 2007, p. 57ff).[11] There is no experience of an aesthetic quality without an emotional interpretant – without being attracted to or repelled by it (cf. C.P. 5.475; Hocutt, 1962, p. 158).

Aesthetics only studies what makes beautiful as it is, and what makes ugly as it is. It does not choose between them as morally preferable. It discusses the relationship between qualities, our relationship towards them, and in that way it establishes what is that which is admirable in itself, what makes it that way, approximating the nature of the *summum bonum* (cf. C.P. 2.199; Potter, 1997, pp. 33–35).

To Peirce, beauty is *a quality.* As such, it is not defined in terms of quantity, forms or representation and is irreducible to the multiplicity of parts. Parret opposes Peirce's views to the classical theories of beauty, like Plato's, and presents his views as "truly contemporary and *avant-garde*", that is, "non-classical, non-cognitivist and non-representational" (Parret, 1994, p. 182).

Beauty is something admirable and is not confined to symmetry and form. But *unbeautiful* is an important aesthetical category, too, because aesthetics is interested in *all* aesthetic goods or qualities. Philosophers have failed to acknowledge the unbeautiful, omitting the most important question of aesthetics: "What is the one quality that is, in its immediate presence, καλός?" (C.P. 2.199).

The term "καλός" is *closer* to what Peirce understands as the subject-matter of aesthetics, dismissing the term "beautiful" as a term that fails to

11 We can observe some interesting similarities and differences between Peirce and Kant. Like Kant, Peirce detaches beauty from any other interest that may belong to other spheres, be it moral or theoretical. Unlike Kant, though, to Peirce, the experience of beauty is not limited to a form of things that is related to the structure of our reasoning. Though systematicity and coordination are part of being beautiful, beauty is not only that. It is an aesthetic quality conveyed by a form, and is an object of admiration. Interestingly, at this point Peirce again stands closer to Kant, than, for example, Hegel, for whom the only subject matter of aesthetics, and the only things that can be labeled as beautiful – are the *works of art* (Cf. G. W. F. Hegel, 1988, pp. 1–3). Unlike Hegel, Peirce and Kant would agree that works of nature too are legitimate objects of aesthetic appreciation. Though Peirce shares some similarities with Kant, their fundamental philosophical stances differ. Peirce's critical stance towards the idea of transcendental philosophy and method, and thus *the idea of a transcendental subject* is what marks a crucial point of divergence between the two philosophers. In addition to that, to Peirce, aesthetics is not only interested in beauty, but all aesthetic qualities, even the qualities of the *unbeautiful.*

acknowledge it (cf. C.P. 2.199ff). Though καλός may be closer, it does not completely convey it. Peirce prefers "καλός" because of its implications for his understanding of an aesthetic quality and that which is admirable in itself.

Καλός conveys some of its important aspects, like unity, purposefulness, and enjoyability. These are the things that make an end admirable: it is purposeful, it is good, and thus lovable (cf. Liszka, 2021, p. 12ff). Since, in Peirce's view, ethics relies on aesthetics in determining the nature of *summum bonum*, that which is in itself admirable and good is not morally determined as such. That is why aesthetic goodness is not defined on the base of moral goodness, but the other way around.

Aesthetics should define what makes an end admirable, and ethics should choose between ends as morally good or bad, that is, to propose goals a man may reasonably choose in different circumstances (cf. Potter, 1997, p. 34). It seems that the understanding of "good" is ontological rather than ethical. It is "good" in terms of making something existent, not because of possessing certain moral qualities. If it acts as an ideal, it is something good, *an end*.

What aesthetics should be is not the theory of fine arts, but of that which is *admirable in itself* (cf. Liszka, 2018, p. 207ff). That which is admirable in itself "without any ulterior reason" is what Peirce calls an *ideal* (cf. C.P. 1.191). An ideal, however, is what we, so to speak, *mimic* in our conduct.

2.2. Aesthetic education: Plato and Peirce

At this point, I would like to shed some light on the difference between Plato and Peirce, to further clarify Peirce's stance on aesthetics. We may argue that both philosophers would agree on the importance of aesthetic education. Plato tackles this problem in *The Republic*, noticing that correct aesthetic education is necessary for any further moral and intellectual improvement (Plato, 1997, pp. 1032–1040; 394e–403c).[12]

Plato criticizes famous poets regarding the way they portray gods and their characteristics (Plato, 1997, pp. 1005–1016ff, 1018–1030; 365eff, 377dff, 379d–383c, 386a–392c). In book II of *The Republic*, Glaucon presents an example of the influence of the poets on people's views. While telling the story of Gyges, Glaucon (Plato, 1997, pp. 1001; 360b-c) presents an image of a powerful man, introducing an interesting observation:

> Now, no one, it seems, would be so incorruptible that he would stay on the path of justice or stay away from other people's property, when he could take whatever he wanted from the marketplace with impunity, go into people's houses and have sex with anyone he wished, kill or release from prison anyone he wished, and do all the other things that would make him *like a god* among humans.[13]

12 More on Plato's views on myths and education see (Deretić, 2014, pp. 233–266).

13 Italics are mine.

The observation "like a god" is important, because it reflects the people's concept of that which is divine. Aside from the problem of the blasphemous depictions of gods, another issue is the practical consequences that these depictions may produce, if accepted. If people see gods not only as supernatural and powerful beings, but also as role-models, then people might start to mimic the behavior of characters so depicted. This blasphemy reflects on people's values. This is just one example of why Plato considered censorship of art necessary.

Plato recognizes that, from an early age, works of art and artistic representations of reality have their fair share in shaping our values. It is why discussions concerning the questions of an ideal state and virtues must include thorough discussions of the role and place of arts and artists in the state. Just as early physical education is necessary for maintaining physical health and healthy habits, so is aesthetic education at an early age necessary to map the further shaping of one's spirit (Plato, 1997, pp. 1032–1040; 394e–403c).

We could say that Peirce, to some extent, would have agreed with Plato in this regard (cf. C.P. 1.592; 2.199). Peirce recognizes that the logically good rests on the morally good, and the latter on the aesthetically good (cf. C.P. 5.131ff). From the point of view of Peirce's aesthetics, however, we could say that the question of aesthetic education cannot be limited to matters related to taste and the influence of art. It is because an artistic preference is already a result of some *previous experiences*. According to Peirce, our aesthetic appreciations are shaped by a more fundamental experience.

Certainly, Peirce's aesthetics would not be limited to a theory of art, its role in a society, and its influence on an individual (cf. C. M. Smith, 1972, pp. 21–29). That would be only a portion of what it might be dealing with. Our aesthetic being is not influenced only by art, but by the qualities of feeling present *in every phenomenon* of possible experience. Rightly addressed, the problem of aesthetic education reveals itself to be the question of how we form and pursue ideals and values that reflect both in ethical and logical conduct.

3. Aesthetics as a normative science

3.1. Feelings and personality

The concept of aesthetics can also be elucidated with the parallels Peirce draws between its subject and the psychological exemplifications of his phenomenological categories (cf. C.P. 8.225–8.256). In discussing the nature of normative sciences, Peirce says that aesthetics is to ethics, as ethics is to logic, the three having their appropriate subject, namely the psychological exemplifications of phenomenological categories of Firstness, Secondness, and Thirdness, in the forms of *feelings*, *action*, and *thought* (ibid).

Peirce's understanding of what aesthetics should be is based on his account on *Firstness*, the concept he saw as a point of divergence with other philosophers; the concept presents his view on the role of the immediate as a part of reality, and experience (cf. C.P. 8.267ff; 8.327ff). He took a critical stance towards the idea of *the given*, and he makes a distinction between the given and *the immediate* (cf. Dewey, 1935, pp. 701–708). That which is immediate cannot be given in the form of an object. Furthermore, that which is immediate cannot be subjective, or objective, since any distinction and relation between subject and object is already a result of a mediation.[14] That which is given has already been mediated by reflection.

However, the immediate, the monadic aspect of reality, presented by the concept of Firstness, is a necessary part of reality and experience, because without it no relation or mediation is possible. The concept of Firstness is one of the fruits of Peirce's attempts to mark the differences between his and, as he called it, the nominalistic metaphysics, which refers to the entire "modern spirit" (cf. C.P. 5.61; 8.267ff; 8.327ff). If feelings are exemplifications of Firstness, then a feeling is the first, or a ground on which the personality operates.

Peirce's claims that he is not well familiar with aesthetics may be interpreted as resting on his convictions on *what must be* its main subject. In this regard, phenomenology provides us with valuable insights. It suggests how we are to understand the relationship between feelings, or the qualities of feeling, to *acting* and *thinking*.

Acting and thinking have been accompanied by a certain quality of feeling (cf. Dewey, 1935, pp. 701–708). This does not mean that they can be *reduced to* feelings. Thought and action have their laws and regularities, but they are not only that. A man is not only acting and thinking, and if the First is exemplified in a quality of feeling, then feelings have an important role in Peirce's understanding of man. One particular passage (C.P. 1.574) sheds some interesting light in this regard:

> It is true that the Germans, who invented the word,[15] and have done the most toward developing the science, limit it to taste, that is, to the action of the *Spieltrieb*[16] from which deep and earnest emotion would seem to be excluded. But

14 Cf. (Hocutt, 1962, p. 158): "In Peirce's terminology, *qualities* are objectively, what *feelings* are subjectively". Also cf. Dewey, 1935, pp. 701–708).

15 Namely, aesthetics.

16 Liszka (2021) draws attention to the importance of Friedrich Schiller's concepts of *lebende Gestalt* (*a living shape*) and of *Spieltrieb* (*the play impulse, play drive*). They

> in the writer's opinion the theory is the same, whether it be a question of forming a taste in bonnets or of a preference between electrocution and decapitation, or between supporting one's family by agriculture or by highway robbery.

This passage aims to explain why the subject of aesthetics has to be broadened. Aesthetics should not be limited to the questions of taste and art, but to examine what values, desires, intentions, and convictions shape our existence and our intellectual efforts to conceptually articulate the world around us, too. The preferences in artistic taste are already *a result*, a judgment, the same as other activities and preferences a man may have. To react to a piece of art is *to already be disposed*, in a way, to do so.

Our fundamental relationship towards reality is aesthetic. If, however, the subject of aesthetics is related to phenomenology, Firstness and qualities of feeling, then it should study not only fine arts and our relationship to them, but that which makes such a relationship possible. As a person with certain dispositions, a man reacts to things in a certain way. It is not likely that two people can have identical aesthetic experiences. Although a piece of art, like a painting or a music composition, convey certain qualities/feelings, they are not encountered in the same way by different people.

We may assume a certain degree of similarity. However, personality is understood as a *teleological harmony in ideas* (C.P. 6.155–6), and it would be difficult to suppose the complete identity of the content of two teleological systems, especially regarding a possible course of action two people may take. We may follow the same ideal/end, but hardly in an identical way.

Regarding the concept of personality, Peirce pays attention to "the saying [...] that man is a bundle of habits" (C.P. 6. 228). In the passage where Peirce is discussing that view, he regards the "bundle" as of secondary importance for our understanding of the nature of personality. If we understand personality as a bundle alone, we would still be missing that which holds this bundle together. In Peirce's words: "[A] bundle of habits would not have the unity of self-consciousness. That unity must be given as a centre for the habits" (C.P. 6.228).

The "centre" that holds the multitude of elements that constitute a person, Peirce had previously mentioned in the same passage, as the "*quale*-consciousness" (cf. Ibid). It is that which makes the unity of con-

---

refer to the harmonious relationship between reason and feeling, which Peirce finds significant and interesting, especially regarding his view on *concrete reasonableness* (cf. Liszka, 2021, p. 12ff, 167, 205; C.P. 5.3). The concept of *Spieltrieb* is also important for Peirce's idea of *musement*, a free activity that lies not only in the base of artistic creativity, but scientific inquiry, too. From the view of Peirce's aesthetics, the concept outlines a relationship towards something, that is free of any other, heterogeneous purpose or motive. On the role of the concept of musement in Peirce's thought see (Cooke, 2018, p. 2ff).

sciousness possible.[17] Peirce argues that the unity of consciousness cannot be understood as a logical, or psychological (physiological), but *metaphysical* (cf. C.P. 6.229):

> The brain shows no central cell. The unity of consciousness is therefore not of physiological origin. It can only be metaphysical. So far as feelings have any continuity, it is the metaphysical nature of feeling to have a *unity*.

Unfortunately, we cannot analyze this passage thoroughly. To do so, we would have to at least pay attention to the concepts of *metaphysical*, *unity*, *consciousness*, *feeling*, and their relationship to each other. That is beyond the scope of this paper. What we may do, though, is notice that Peirce avoids using conceptual categories of psychology and physiology when discussing the nature of *personality*.

A unity of a self-conscious personality, or a teleological harmony of ideas, is understood as a *qualitative* unity. In that manner, Peirce also describes personality as a *living feeling* (C.P. 6.155). The view of personality in terms of teleological harmony in ideas reveals an interesting relationship between that and the concepts of "teleology" and "feeling" (cf. C.P. 6.156).

Personality, Peirce notices, is "a phenomenon which is remarkably prominent to our own consciousness" (C.P. 6.155). He is not interested, though, in discussing its neurophysiology, but the general idea (cf. Reynolds, 2002, 63ff). Generally speaking, personality is a "coordination or connection of ideas" (C.P. 6.155). However, that connection is "itself a general idea, and that general idea is a living feeling" (ibid). Ontologically and epistemologically, it is not a thing among other things, a finite object that can be completely defined, or "apprehended in an instant". Rather:

> It has to be lived in time: nor can any finite time embrace it in all its fullness. Yet in each infinitesimal interval it is present and living, though specially colored by the immediate feeling of that moment. (C.P. 6.155)

The notion of a feeling that colors personality is interesting. A person is not something that can be fully described in logical, psychological, or physiological terms. As a developmental teleological system, it is determined by the *qualities of feeling*.

The formation of an ideal, as that which we follow and which guides our activities, is tied to a feeling that colors personality. The ideal is what

---

17 Peirce (C.P. 6.226) contrasts this notion with Kant's understanding of the synthetic activities of intellect: "and the various synthetical unities which Kant attributes to the different operations of the mind, as well as the unity of logical consistency, or *specific* unity, and also the unity of the individual object, all these unities originate, not in the operations of the intellect, but in the *quale*-consciousness upon which the intellect operates".

allows for our behavior to be described as goal-oriented. It is that which recommends itself as desirable, and it is deeply related to a quality of feeling. As we have already said, *the First* is to be understood as both a quality of feeling which *colors our personality* (C.P. 6.155), and as something that determines *an ideal* to pursue (cf. Parret, 1994, pp. 179–180).

It is hardly avoidable to consider the role of feelings when discussing our goal-oriented activities. Furthermore, the topic is hard to avoid regarding the issue of the *unity of personality*. That unity is *qualitative*; it is *indivisible* and *irreducible* to its parts. Approaching this issue from aesthetic point of view, we can notice some similarities to Peirce's understanding of the relationship between beauty and symmetry. Just as non-beautiful parts only make beauty to appear, but are not capable on their own to generate beauty and to be equated with it, so the aforementioned "bundle of habits" is only what a unity of personality *appears through*.

That unity has been determined by something pre-reflexive, and our habits and thinking, our complete teleological *modus operandi* may be deeply influenced by it. Following that notion, we may say that to further understand what a man is, is to understand the nature of the aesthetic appreciations.

A man is the thought (cf. C.P. 5.314–315), but a thought, as an instance of the Third, is impossible without the Second and the First. They are not identical, of course. But a philosophic-anthropological view on what a man is, must take into consideration all the categories. Pierce's view of man is determined by his phenomenology, and the understanding of the nature of aesthetic experience.

4. Concluding remarks

Peirce was neglecting aesthetics throughout his career, a fact he was well aware of. Yet, his relationship to it is more complex than that. His early encounter with philosophy was through studying Schiller's *Letters*, whose ideas made a lasting influence on Peirce's thought. However, Peirce developed original notions on aesthetic experience, and aesthetics, that rest largely on his phenomenology and the theory of categories, which is probably why he could not develop the notions the way he did earlier. The importance of aesthetics rests on the relevance of the concept of Firstness, given that it is supposed to study some of its exemplifications. That is why he contrasts his views with the traditional. The main flaw of traditional aesthetics is in not recognizing that its alleged subject-matter was secondary to something more important. Because of that, aesthetics limits itself to the matters of taste, fine arts and various notions of the beautiful.

According to Peirce, as a normative science, aesthetics should study various aesthetic qualities, suggesting how we ought to reflect on them. Since it pertains to that which is *admirable*, it pertains to that which we appreciate for its own sake. Consequently, it should outline what it is in a thing that we find cherishable, desirable, admirable, which is an issue that falls outside the scope of ethics and logic.

Aesthetics should observe the mechanics of forming and following an ideal. A man, as a teleological system, cannot do without having different ideals. An ideal is that which allows our behavior to be described as purposeful and goal-oriented. Aesthetics recognizes the close relationship between the concepts of the highest good, and aesthetical qualities and experiences. That is why an account on the teleological form of our conduct would require a theory like that. Ultimately, such a theory would provide a strong aid to a philosophic-anthropological account on what a man is.

Bibliography

Anđelković, M. (2006). Pers o percepciji i abdukciji. *Theoria, Beograd*, *48*(4), 7–16.

Aristotle. (2009). *The Nicomachean Ethics* (Oxford World's Classics; D. Ross, Trans.). Oxford New York: Oxford University Press.

Aristotle. (2018). *Physics* (C. D. C. Reeve, Trans.). Indianapolis/Cambridge: Hackett Publishing Company, Inc.

Atkinson, N. P. (1982). My Life written for the Class-Book. In M. H. Fisch, C. J. W. Kloesel, E. C. Moore, D. D. Roberts, & L. A. Ziegler (Eds.), *Writings of Charles S. Peirce: A Chronological Edition, Volume 1* (pp. 1–3). Indiana University Press. JSTOR.

Cooke, E. F. (2018). Peirce on Musement. *European Journal of Pragmatism and American Philosophy*, *X*(2). https://doi.org/10.4000/ejpap.1370

Deretić, I. (2014). *Platonova filozofska mitologija*. Beograd: Zavod za udžbenike i nastavna sredstva.

Dewey, J. (1935). Peirce's Theory of Quality. *The Journal of Philosophy*, *32*(26), 701–708. https://doi.org/10.2307/2016281

Frank, M. (2008). *The Philosophical Foundations of Early German Romanticism* (E. Millán-Zaibert, Trans.). New York: State University of New York Press.

Gorodeisky, K. (2016). 19th Century Romantic Aesthetics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/fall2016/entries/aesthetics-19th-romantic/

Hegel, G. W. F. (1988). *Aesthetics: Lectures on Fine Art* (Vol. 1; T. M. Knox, Trans.). Oxford: Clarendon Press.

Hegel, Georg Wilchelm Friedrich. (1977). *Phenomenology of Spirit* (A. V. Miller, Trans.). Oxford, New York, Toronto, Melbourne: Oxford University Press.

Hocutt, M. O. (1962). The Logical Foundations of Peirce's Aesthetics. *The Journal of Aesthetics and Art Criticism*, *21*(2), 157–166. https://doi.org/10.2307/427188

Kant, I. (2007). *Critique of Judgement* (J. C. Meredith, Trans.). Oxford, New York: Oxford University Press.

Liszka, J. J. (2018). Peirce's Esthetics As a Science Of Ideal Ends. *Cognitio: Revista de Filosofia*, *18*(2), 205–229. https://doi.org/10.23925/2316–5278.2017v18i2p205–229

Liszka, J. J. (2021). *Charles Peirce on Ethics, Esthetics and the Normative Sciences*. New York: Routledge Taylor & Francis Group.

Milisavljević, V. (2006). *Identitet i refleksija: Problema samosvesti u Hegelovoj filozofiji*. Beograd: Zavod za udžbenike i nastavna sredstva.

Parret, H. (1994). Peircean Fragments on the Aesthetic Experience. In H. Parret (Ed.), *Peirce and Value Theory: On Peircian Ethics and Aesthetics* (pp. 179–190). John Benjamins Publishing Company.

Peirce, C. S. (1982). My Life written for the Class-Book. In M. H. Fisch, C. J. W. Kloesel, E. C. Moore, D. D. Roberts, & L. A. Ziegler (Eds.), *Writings of Charles S. Peirce: A Chronological Edition, Volume 1* (pp. 1–3). Indiana University Press.

Petrović, A. M. (2016). ΦΥΣΙΣ i ΚΟΣΜΟΣ: priroda i svet kod antičkih mislilaca sa optikom Kantove tehnike prirode. Novi Sad: Zavod za kulturu Vojvodine.

Plato. (1997). Republic. In J. M. Cooper (Ed.), & G. M. A. Grube (Trans.), *Plato: Complete works* (pp. 971–1223). Indianapolis, Ind: Hackett Publishing.

Plotinus. (2018). *The Enneads* (L. P. Gerson, Ed.; G. Boys-Stones, J. M. Dillon, L. P. Gerson, R. A. H. King, A. Smith, & J. Wilberding, Trans.). Cambridge: Cambridge University Press.

Potter, V. G. (1997). *Charles S. Peirce: On Norms & Ideals*. New York: Fordham University Press.

Rescher, N. (1987). Aporetic Method in Philosophy. *The Review of Metaphysics*, *41*(2), 283–297.

Reynolds, A. (2002). *Peirce's Scientific Metaphysics. The Philosophy of Chance, Law and Evolution*. Nashville: Vanderbilt University Press.

Shields, C. (2023). Aristotle. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2023/entries/aristotle/

Smiljanić, D. (2020). *Biće od trske: Antropologija u sinestetičkom pogledu*. Novi Sad: Adresa.

Smith, C. M. (1972). The Aesthetics of Charles S. Peirce. *The Journal of Aesthetics and Art Criticism*, *31*(1), 21–29. https://doi.org/10.2307/429607

https://doi.org/10.18485/baf.2025.2.ch15

Natalia Tomashpolskaia

LUDWIG WITTGENSTEIN'S CRITIQUE OF THE DISPOSITIONAL THEORY OF VALUES

Abstract: This paper discusses Ludwig Wittgenstein's critique of the dispositional theory of values and argues that the later Wittgenstein was not fully a conventionalist. In his later writings, notes, and conversations, Wittgenstein engaged with ethics, aesthetics, and religion, in ways that resist reductive classification within analytic philosophy. While he recognised the importance of *as-if* conventions, language-games, learned behavioural patterns, the plurality of world-views and ethical systems, the various expressions of religious experience, and subjective attitudes in shaping personal values, he also defended the existence of ethical and aesthetic values that are not merely conventional.

Key words: Ludwig Wittgenstein; Values; Dispositions; Conventions; Ethics

Introduction

Ludwig Wittgenstein's reflections on ethics, aesthetics, and values present a challenge to the dispositional theory of values, which ties values to subjective attitudes, social conventions, or behavioural tendencies. While Wittgenstein's later philosophy is frequently associated with the concepts of language-games and conventionalism, his remarks on values and ethics offer a more complex account. He recognised the role of conventions, calling them "as-if conventions", learned social practices, and subjective preferences in shaping how human beings understand values. Yet, at the same time, he argued for the existence of values that transcend mere conventions – values that can be described as higher values in a non-conventional sense.

This article offers an analysis of Wittgenstein's philosophical works from his return to Cambridge in 1929 and the beginning of his so-called later or second period, as well as personal recollections and notes from close associates such as Norman Malcolm, Rush Rhees, M. O'Connor

Drury, Oets K. Bouwsma, and G.E.M. Anscombe. It explores Wittgenstein's critique of the dispositional theory of values, focusing on his understanding of *as-if* conventions, his distinction between aesthetics and psychology, his conception of ethics as fundamentally distinct from what he understood as ethical systems – moral customs or sociological descriptions – and his rejection of the idea that values (e.g., good) can be reduced to subjective or cultural preferences.

Discussing the influence of the Kantian critical philosophy on the development of new philosophical schools in the twentieth century, R. Hanna (2008, p. 158) explains the origin of the analytic philosophy during the late nineteenth and early twentieth centuries (mid-1930s) as a reaction against both neo-Kantianism and neo-Hegelianism, which pushed scholars toward Platonism and radical realism. At the same time, the antimetaphysical position of logical positivism pushed analytic philosophy in the opposite direction, towards conventionalism and anti-realism. According to Hanna (2008, p. 158), this inherent tension and conflict "between Platonism and realism on the one side, and conventionalism and anti-realism, on the other," formed the foundation for the later development of key concepts in the field, such as the *analytic proposition* (a necessary *a priori* truth based on logical laws and logical definitions) and the method of *analysis* (the process by which such propositions are understood). I propose that Wittgenstein's later work provides a departure from these dichotomies central to analytic philosophy. Wittgenstein rejected both metaphysical realism and anti-realism as philosophical and grammatical confusions. On the one hand, he acknowledged conventionalism of the intersubjective level of language-games and human interactions; however, Wittgenstein called social conventions "*as-if*" conventions. These conventions do not imply the agreement between human subjects, but the agreement of an individual to follow or not to follow yet established social norms and rules. And, on the other hand, Wittgenstein in his later words (from 1929 until 1953) still adheres to the special understanding of ethics and aesthetics as realms that transcendent ordinary language conventionalism, both societal and between individuals. Saying that the Good is divine and supernatural, that the Beautiful is never agreeable, but aesthetic reasons let us get closer to an ideal or farther from it, and that the only rule in aesthetics is the "rule of harmony" but not subjective tastes, Wittgenstein challenges further scholars leaving room for his later work to be accused even of Platonism.

For instance, Michael Dummett (1978) interpreted Wittgenstein's later philosophy, particularly in mathematics and logic, as a form of *radical conventionalism*. Although Dummett's interpretation is deeply rooted in the texts, it focused on logic and mathematics and did not discuss aesthetics and ethics, and it has been largely rejected for philosophical reasons (see McDowell 1984; Wright 1980). However, the rejection of Dummett's interpretation often seems to rely on a form of *argumentum ad lapidem*[1] where radical conventionalism is dismissed as evidently untenable, and, as a result, it is assumed that Wittgenstein could not have held such a view. Conventionalism, broadly understood, maintains that the truth-value of propositions within a given domain is determined by linguistic conventions (see Quine 1966; Glock 2008; Warren 2015; Topey 2019), or, in other words, that such propositions are true in virtue of meaning (Glock 2003). Dummett's articulation of radical conventionalism arises from his critique of more moderate variants. Building on Dummett's work, Ásgeir Berg (2024) defends the view that Wittgenstein held a radical conventionalist position, where mathematical truths are determined by our practices rather than pre-established rules.

However, a number of scholars have focused on and further enriched the discussion of Wittgenstein's understanding of ethics and its complex relationship to conventionalism. G. E. M. Anscombe, strongly influenced by her close engagement with Wittgenstein, investigated the significance of intention and action in moral life in her book *Intention* (1958). Peter Winch (1958) extended Wittgenstein's ideas into the domain of the social sciences, showing how moral reasoning is embedded within distinct forms of life. Other scholars, such as Peg O'Connor and Anne-Marie Søndergaard Christensen, influenced by late Wittgenstein, examine how ethical practices emerge from communal norms while retaining a critical independence from them. O'Connor (2000) analyses moral responsibility within the structures of oppression, whereas Christensen (2007) explores the possibility of objectivity in ethics without appealing to universal standards. Hämäläinen (2015), in turn, highlights how moral frameworks shift historically, working within Wittgenstein's understanding of ethical systems and plurality of world views. Raatzsch (2010), focusing on the intersections of language, life, and ethical responsiveness, offers a detailed study of moral self-understanding. Even R. M. Hare's (1952) theory of universal prescriptivism echoes certain Wittgensteinian concerns, particularly regarding the logical and normative structure of moral language. Among the most recent works, Iczkovits (2012) and Wisnewski (2007) have argued that Wittgenstein offers a framework for understanding values as expressions of a worldview that resists both naturalistic reduction and relativistic pluralism. Iczkovits emphasises Wittgenstein's idea of "seeing the world rightly" as a non-empirical ethical stance irreducible to so-

1 "Appeal to the stone" represents a logical fallacy.

cial norms or psychological states. Wisnewski re-examines historical-ethical theories through a Wittgensteinian lens, suggesting that ethics should clarify what is involved in our complex "form of life" rather than dictate prescriptive norms.

Returning to the interpretation that Wittgenstein's later work only partially endorses conventionalism, this paper argues that his reflections on ethics and aesthetics pose a challenge to both naturalistic and relativistic tendencies. Following Richard Amesbury's (2020), Wittgenstein's ethics is approached not as a body of rules, but as a reflective and critical practice integral to human life; one "cannot lead people to what is good", ethical values are "supernatural", they cannot be written down as a list of norms and regulations. Wittgenstein strongly criticised any dogmatism but, at the same time, he advocated the plurality of world views, ethical systems as habitual customs and norms within a form of life of a cultural-historical society, and the plurality of the expression of religious belief, symbolism, rituals. Here we should not be confused by an evident contradiction between the understanding of ethics. For later Wittgenstein, the domain of ethical and aesthetical as supernatural and inexpressible is still relevant and he continues talking about it in his later works following the ideas that were best expressed in his *Lecture on Ethics*. He argues that the good is supernatural and that aesthetics bears no relation to subjective tastes. At the same time, however, in his later works Wittgenstein discussed ethical propositions as expressions of subjective judgement, shaped by learned linguistic and cultural–societal practices. These so-called ethical judgements, along with judgements of taste, are subjective and epistemologically distinct from the ethical – which is described as "divine" and "supernatural" – and the aesthetic, which follows the rules of harmony and aspires toward the ideal.

Conventions

Wittgenstein argues that viewing language merely as a lens or as a simple tool for communication is an overly simplistic and limited way of understanding its role and nature. Moreover, according to the recollections of Wittgenstein's friends and colleagues and his diary notes it would be unjust to employ terms such as *conventionalism* or any other such "-isms" in reference to Wittgenstein's thought, out of respect for his well-known personal aversion to such classifications. Wittgenstein himself found the application of philosophical labels to his work both unpleasant and misleading. He was firmly opposed to belonging to any particular school of thought or creating a new one. I shall argue that the later Wittgenstein was not a conventionalist philosopher in a straightforward sense.

For instance, in the *Yellow Book*, §12, – one of earliest composed text of the beginning of his late period – he writes that even "what is possible or impossible is an arbitrary matter" (Wittgenstein 2001, p. 64). He goes on to state that "we could make it a rule, for example, that "green and yellow can be in the same place at the same time" is to make sense." In §13 he notes, "we tend to think of a possibility as something in nature, something we are able to imagine" and that "when one talks of possibility, one is making use of picture"; indeed, "we are tempted to say of the possibility that it is potentially present" (Wittgenstein 2001, p. 64). In these lecture notes, §18, he further asserts that even "The laws of logic, e.g., excluded middle and contradiction, are arbitrary" (Wittgenstein 2001, p. 71). He adds that "this statement is a bit repulsive but nevertheless true". That these laws are arbitrary is of particular significance, for instance, in mathematics. Here, the contradiction is of the form *p* and *not-p* (Wittgenstein 2001, p. 72). Wittgenstein observes, "To forbid its occurrence is to adopt one system of expression" – a system that differs from our ordinary one, in which such contradictions are excluded. He explains, "if we say the thing can"t at the same time be both red and not-red, we mean that in our system we have not given this any meaning. An adopted system of expression is like an adopted measuring rot."

The key issue, for Wittgenstein, lies in application – we cannot meaningfully discuss a system of signs without applying to their use. The application of the law of contradiction is not independent of the law itself. As he puts it in the *Big Typescript*, "In so far as this series of patterns is a series of signs, a new sign enters language with each new pattern" (Wittgenstein, 2005, p. 328e).

In his 1934–35 lectures, Wittgenstein provided a clear explanation of how he understood the concept of a convention. "By a convention I mean that the use of a sign is in accordance with language habits and training" (Wittgenstein, 2001, p. 89). A convention can form part of a chain: "a convention is established by saying something in words," or more generally, it is "something laid down by a sign". One might say that "signs play the role they do because of certain habitual ways of acting" (Wittgenstein, 2001, p. 90). In these lectures (Lecture XIV), Wittgenstein attempted to elucidate the ambiguity inherent in our understanding of convention. For instance, to regard the expression "2+2=4" as a convention is misleading, although this equation "might originally have been the result of one" (Wittgenstein, 2001, pp. 156–7). This situation is analogous to the so-called *Social Contract Theory* – though no historical contract ever occurred, in Wittgenstein's words, "it is as if such a contract has been made". In *Philosophical Grammar* (Wittgenstein, 1974, p. 30) §138, Wittgenstein writes that "grammar consists of conventions," and further clarifies: "But it is the connection and not the effect which determines the meaning." He adds: "an explanation of meaning is not an empirical proposition and not a causal explanation, but a rule, a convention" (Wittgenstein, 1974, p. 68). As he puts it in the *Big Typescript*: "'Contrat sociale' [sic] – here too, no actual contract was ever concluded; but the situation is more or less similar, analogous, to the one we'd be in, if [...] And there's much to be gained in viewing it in terms of such a contract" (Wittgenstein, 2005, p. 151e).

The same applies to the equation "2+2=4"– again, it is as if "the convention had been made." Wittgenstein notes that "2+2=4" is an "instrument", comparable to a table or any other object. Thus, in real life and in ordinary language, people more often operate with *as if* conventions than with explicitly stated rules governing meaning. For example, Wittgenstein observes that "the assertion of numerical equality is a proposition of grammar, and says nothing about reality" (Wittgenstein, 2001, p. 161). Similarly, "To say that 10×10 = 2×10 [...] is to assert a proposition of grammar; it is not about the world". The same explanation applies for geometrical lines – "the statement about the geometrical lines does not say anything about reality". Nevertheless, Wittgenstein notes that "we might have different conventions for counting". For instance, "we might say the pentagram has six vertices because we had decided to count one vertex twice over" (Wittgenstein, 2001, p. 172). Yet, people tend to believe that a pentagram *must* have five vertices by its very nature – that this is "in no way dependent on a convention", and that the image of a pentagram demonstrates "something in its essence". Wittgenstein warns: "this is a dangerous trap" (Wittgenstein, 2001, pp. 172–3). In the lectures recorded by Moore, Wittgenstein remarked that we understand something "by means of convention" (Wittgenstein, 2016, p. 11). These conventions, he explained, are taught through the use of language: "these conventions are made by giving a verification of the proposition. This establishes a connection between language & your expectations." He further stated: "You understand" means a "sentence arouses in you something related to reality, in same way as expectation to reality." However, he later clarified that "grammar isn't mere convention". "When we talk of rules of grammar, we have in mind that words are used in particular ways in significant sentences" (Wittgenstein, 2016, p. 28).

In the *Big Typescript* (Wittgenstein, 2005, p. 145e), Wittgenstein writes that when "looking at language as a social institution that is subject to certain rules", we cannot justify these rules; rather, we can only "describe them as a game that people play". He continues: "Grammatical rules are not those (it goes without saying: empirical) rules in accordance with which language has to be constructed to fulfil its purpose. In order to have a particular effect." Grammatical rules are thus merely descriptions of how language operates. Grammar, he says, describes "the game of language, the linguistic actions". In *Philosophical Grammar* §134 (Wittgenstein, 1974, p. 186) Wittgenstein addresses a crucial distinction between what he calls rules of *representation* and *conventions*. He argues that rules of representation are not mere conventions if their validity can be justified by their ability to produce representations that correspond to or align with reality. In other words, the point here is that if the application of a rule of representation ensures that a representation accurately reflects reality, the justification for the rule cannot simply be that it conforms to some external convention or arbitrary standard. Wittgenstein's main point is that the justification of grammatical or representational rules cannot rest solely on the fact that their application results in true or accurate representations. To use a rule simply because it produces an accurate representation of the world would reduce the rule to something like a technical convention, where the application of the rule becomes a matter of pragmatic utility rather than a matter of meaning or necessity within the language game itself. Further, Wittgenstein adds that such a justification would itself have to describe the content of the representation, thus leading to a circular problem. To justify a rule by showing that it leads to a representation that agrees with reality would require describing what is represented, which means we would end up needing a further rule to explain what is represented by the representation itself. This creates an infinite regress, where the justification for each rule would ultimately require the application of yet another rule, and so on. This circularity demonstrates that the validity of grammatical or representational rules cannot be grounded solely in their correspondence to an external reality; rather, they must be understood as part of the framework of language, which is grounded in the practices and forms of life that give it meaning. In essence, Wittgenstein is emphasising that grammar, and the rules governing language, cannot be reduced to mere conventions based on how well they map onto the world. Grammar, rather, must be seen as an internal structure of the language, a set of rules that govern the way meaning is formed and communicated, independent of any direct appeal to external reality. The justification for these rules is not empirical or based on correspondence with the world, but inherent in the way language operates within a given form of life. In other words, for Wittgenstein conventions are limited by the grammar of language. Grammatical rules unlike rules of a language-game (that could be conventional) have an "application to reality".

Michael Dummett argued that Wittgenstein had fallen into a "full-blooded conventionalism". However, Wittgenstein's position is diametrically opposed to that of Austin, whose conventionalism is best expressed in his own words: "We are absolutely free to appoint any symbol to describe any situations, as far as being merely true goes" (Austin 1950, p. 118). For Wittgenstein, conventions are constrained by the grammar of language. Unlike the rules of a language-game – which may indeed be conventional – grammatical rules possess an "application to reality". Wittgenstein's "conventionalism" is therefore not reducible to a simple agreement or convention among human beings. In *Zettel* he explains:

> How could human behaviour be described? Surely only by sketching the actions of a variety of humans, as they are all mixed up together. What determines our judgement, our concepts and reactions, is not what *one* man is doing *now*, an individual action, but the whole hurly-burly of human actions, the background against which we see any action. (Wittgenstein, 1970, p. 99e)

In his notes for lectures (Wittgenstein, 1968, p. 285), Wittgenstein considers the example of blind people: "Now whom shall we call blind?" He answers: "A certain kind of behaviour. And if the person behaves in that particular way, we not only call him blind but teach him to call himself blind." This idea is of great importance, as it highlights that a person is taught to identify with a particular label and to behave in accordance with it. It forms part of social and language-games: to conform to a specific set of behavioural rules appropriate to a social role recognised within a given culture. This framework applies to a wide range of social roles and forms of behaviour – for example, lying, or saying "I have pain".

> The word "lying" was taught us in a particular way in which it was fastened to a certain behavior, to the use of certain expressions under certain circumstances. Then we use it, saying that we have been lying, when our behavior was not like the one which first constituted the meaning. (Wittgenstein, 1968, p. 295)
>
> I have taught the child to use the expression "I have toothache" under certain circumstances. And now it uses the words under these circumstances. (Wittgenstein, 1968, p. 296)

Wittgenstein repeatedly emphasises that ordinary human life is exceedingly complex, likening it to a net consisting of interwoven threads. Yet, as he notes: "Seeing life as a weave, this pattern (pretence, say) is not always complete and is varied in multiplicity of ways" (Wittgenstein, 1968, p. 296). Indeed, we human beings "in our conceptual world keep on seeing the same," (we were taught to see the same) "recurring with variations".

This is how our concepts "work", they are "not for use on a single occasion [*einmaligen Gebrauch*]". One conceptual pattern, therefore, "is interwoven with many others" (Wittgenstein, 1970, p. 100e).

In 1950, Wittgenstein articulates this point clearer: "Practice gives the words their sense" [*Die Praxis gibt den Worten ihren Sinn*] (Wittgenstein, 1980, p. 85e). Accordingly, the key to understanding "agreement" lies in practice. "In fact, if he is to play a language-game, the possibility of this will depend upon his own and other people's reactions. The game depends upon the agreement of these reactions [...]" (Wittgenstein, 1968, p. 283). In *Culture and Value* (Wittgenstein, 1980, p. 40e), Wittgenstein again addresses conventionalism in ordinary life, referring to composers who must choose and adopt a concrete form of counterpoint in line with their own propensities: "He may have his upon a conventionally acceptable attitude and yet still feel that it is not properly his." Similarly, in the *Lectures on Religious Belief*, he speaks of the "connection of convention" (Wittgenstein, 1967, p. 68), providing an example:

> If I said "My brother is in America" – I could imagine there being rays projecting from my words to my brother in America. But what if my brother isn't in America? – then the rays don't hit anything. (Wittgenstein, 1967, p. 67)
>
> The most important point is this – if you talk of painting, etc. your idea is that the connection exists *now*, so that it seem [*sic*] as though as long as I do this thinking, this connection exists.
>
> Whereas, if we said it is a connection of convention, there would be no point in saying it exists while we think. There is a connection by convention. – What do we mean? – This connection refers to events happening at various times. Most of all, it refers to a technique. (Wittgenstein, 1967, p. 68)

In the early 1932–1935 Cambridge Lectures (Wittgenstein, 2001, p. 28), Wittgenstein emphasised, "What is taken as a reason for belief in a theory is thus not a matter of experience but a matter of convention." However, mathematics, for instance, is not a game. "Arithmetic isn't a game, it wouldn't occur to anyone to include arithmetic in a list of games played by human beings" (Wittgenstein, 1974, p. 293). It is not a game because "In arithmetic we cannot make preparations for a grammatical application" (Wittgenstein, 1974, p. 310). While "we can say: arithmetic is its own application. The calculus is its own application." "What is incorrect is the idea that the application of a calculus in the grammar of real language correlates it to a reality or gives it a reality that it did not have before" (Wittgenstein, 1974, p. 311).

Realm of the Non-Conventional in the Later Period

Wittgenstein left room for certain areas where we cannot speak as though conventions and intentional agreements are involved. This brings us to the notion of the "mystical", which lies beyond any such conventions. I will argue that the later Wittgenstein did not entirely dismiss all "mystical" views. In his later notes, remarks, and, especially, in his conversations with friends, we find intriguing passages on ethics, aesthetics, religion, and consciousness that do not neatly fit into the framework of Wittgenstein as merely an analytic philosopher or a conventionalist. In a note dated 15.11.1929 published then in *Culture and Value*, Wittgenstein wrote a famous enigmatic passage introducing his views on the so-called higher values: "What is good is also divine [...]. You cannot lead people to the good; you can only lead them to some place or other; the good lies outside the space of facts" (MS 107 196 in Wittgenstein, 1980, p. 3e).

Wittgenstein did not reject or dismiss all ethical, esthetical and religious reflections in his later philosophy, especially in his personal life, as recorded in his diaries and in conversations with friends. For instance, regarding aesthetics he argued, ""Beautiful" ≠ "agreeable"" (Wittgenstein, 2016, p. 346). Aesthetics, he maintained, is not a matter of mere taste: "The question of Aesthetics is not: Do you like it? But, if you do, why do you?" (Wittgenstein, 2016, p. 346). And he added, "If ever we come to: I like this; I don"t, there is an end of Aesthetics; & then comes psychology" (Wittgenstein, 2016, p. 347). In other words, judgments about what pleases us belong properly to psychology, not to aesthetics. He further observed that "rules of harmony" "always presuppose that we shall understand that obeying these rules makes things more agreeable". For Wittgenstein, aesthetics is categorically distinct from psychology: "I think one thing is clearer: one could sum up: Aesthetic reasons are given in the form: getting nearer to an ideal or farther from it. Whereas Psychology gives causes why people have an ideal" (Wittgenstein 2016, p. 355). Regarding ethics, Wittgenstein differentiated between *ethical systems* as belief systems of morals and customs specific to a particular historical-cultural society, and Ethics as a realm of the "mystical".

Ethics vs. Plurality of Ethical Systems, World-Views and Religious Experience

O'C. Drury, in his recollections of conversations with Wittgenstein, argued that the key to understanding Wittgenstein's thought lies in recognising the central role of *ethics* in the whole of his thought since the very early ideas in the *Notebooks 1914–1916* until his latest writings. In 1948 Wittgenstein said to O'C. Drury, "My fundamental ideas came to me very early in life" (Drury 1984, p. 158). Wittgenstein explained in a letter to Ludwig von Ficker that the fundamental idea of the *Tractatus* was to draw

> limits to the sphere of ethics from the inside, as it were, and I am convinced that this is the ONLY *rigorous* way of drawing those limits. In short, I believe that where *many* others today are just gassing, I have managed in my book to put everything firmly into place by being silent about it. (Wittgenstein, 1979, cited after Rhees 1984, p. 81).[2]

Therefore, his initial idea since *Tractatus* was to distinguish clearly between what we truly know – and hence can speak of with confidence – and what we do not know, refraining from saying more than we know. In the *Blue Book*, written in the beginning of his later period, Wittgenstein reiterates this idea again: "The difficulty in philosophy is to say no more than we know" (Wittgenstein, 1998, p. 44). O'C. Drury maintained that Wittgenstein's aim was precisely to draw rigorous limits to language so as to foreground the specific demands of ethics – not to say more than we really know. He writes:

> Now I am going to venture to state that *all the subsequent writings continue this fundamental idea*. They all point to an ethical dimension. And they do this by a rigorous drawing of the limits of language so that the ethical is put firmly into place. This limitation has to be done from the inside so that whereas nothing is *said* about the ethical it is shown by the rigor of the thinking. (Drury, 1996, p. 81)

O'C. Drury (1996, p. 82) continued: "I believe that difficulty that should be found in understanding Wittgenstein's writing is not merely an intellectual difficulty but an ethical demand. The simple demand that we should at all times and in all places say no more than we really know."

The following passages from Wittgenstein's *Lecture on Ethics*[3] highlight his idea that ethics transcends the limits of language and factual description.

> I can only describe my feeling by the metaphor, that, if a man could write a book on Ethics, which really was a book on Ethics, this book would, with an explosion destroy all the other books in the world. Our words used as we use them in science, are vessels capable only of containing and conveying meaning and sense, *natural* meaning and sense. Ethics, if it is anything, is

2 "Rigorous" for Wittgenstein is not a synonym for "strict" but "'Rigorous' means: clear" (Wittgenstein, 1974, p. 454).

3 Ludwig Wittgenstein delivered a lecture on ethics in Cambridge on 17 November 1929 just after his return to the university.

> supernatural and our words will only express facts; as a teacup will only hold a teacup full of water [even] if I were to pour out a gallon over it. I said that so far as facts and propositions are concerned there is only relative value and relative good, right, etc. (Wittgenstein, 1993, p. 40)

Here, Wittgenstein implies that ethical truths are of an entirely different order, one that transcends the domain of natural language, which is suited only for conveying facts and empirical sense. He compares words to vessels – scientific language can only hold finite meaning, much like a teacup holding water. If ethical meaning were to be poured into these words, their limited capacity would leave them wholly inadequate. For Wittgenstein, facts and propositions can only speak of *relative* value or *relative* good, not of the absolute, which belongs to the ethical or supernatural realm. Thus, he continues, the tendency of people to talk about Ethics and Religion was, in fact, an attempt to "run against the boundaries of language." However, this "running against" is "absolutely hopeless". As well as any description of absolute value is impossible and hopeless, either. This is because any such description, by virtue of being significant (i.e., meaningful within language), would necessarily fail to express what is absolute, which lies beyond linguistic representation. For Wittgenstein, absolute value is ineffable, it cannot be captured or confined within the boundaries of meaningful propositions. This unique supernatural essence of ethics leads to the understanding that a "certain characteristic misuse of our language runs through *all* ethical and religious expressions" (Wittgenstein, 1993, p. 42). When words are used in ethical or religious contexts, their meanings are not identical but analogous to their usage in ordinary contexts. For instance, "all religious terms seem in this sense to be used as similes or allegorically."

In a conversation with Rhees on 2 September 1945, Wittgenstein noted that "*the Sitten und Gebräuche* (morals and customs) of various tribes" *are not* and *would not* be ethical. Describing the morals, customs, habits and cultural behaviour of nations is not the same as studying rules and laws. Accordingly, "A rule is neither a command nor order – because there is no one that gives the order – nor is it an empirical statement of how the majority of people behave" (Citron, 2015, p. 27). For Wittgenstein various uses of a rule are determined by grammar. He remarked that ethics has "special features", as it generally lacks proof (Citron, 2015, p. 28). Further, Wittgenstein introduced the idea of the plurality of *ethical systems* distinguishing them from ethics and understood as conventional systems of shared morals and customs within societies.

> If anyone says that something is good, he is making a judgement of value. If I decide that a certain ethical judgement is true – or that a certain system

> of ethics is the right one – then I am also making a judgement of value. In other words, I should be adopting that system of ethics, or making the same ethical judgement. (Citron, 2015, p. 29)

For Wittgenstein, *ethical systems* can be understood by analogy with language-games, where an individual adopts a particular system with its corresponding statements and rules. However, no system is inherently superior to another; we cannot claim that only one system is the *right* one. To do so would require adopting "certain ethical criteria". Wittgenstein noted the possibility of "logical criticism" to rule out certain ethical systems as incoherent but acknowledged that "this would need further examination".

Wittgenstein also addressed the plurality of *world-views*, maintaining that no single world-view is better than another; all are equally significant. "One could say 'every view has its charm', but that would be false. The correct thing to say is that every view is significant for the one who sees it as significant (but that does not mean, sees it other than it is). Indeed, in this sense, every view is equally significant" (Wittgenstein, 1993, p. 135). When someone asserts that one ethical system is *more right* than another from their point of view, Wittgenstein argued, this reflects only a subjective attitude – "that each judges as he does." He criticised both evaluative statements and the role of reason in ethics: "If you simply take the expression of the judgement – say 'ah', together with a facial expression, — this might be the same for an excellent salad, a great painting or a noble action" (Citron, 2015, p. 30). And in *On Certainty* §128–129 Wittgenstein reflected on the formation of judgements: "From a child up I learnt to judge like this. *This is* judging. This is how I learned to judge; *this* I got to know *as* judgement" (Wittgenstein, 1969, p. 19e). In the next paragraphs in *On Certainty* Wittgenstein challenges the assumption that *experience* serves as the foundation for all forms of judgement. In remark 131, he states unequivocally: "No, experience is not the ground for our game of judging. Nor is its outstanding success." Here, Wittgenstein emphasises that our systems of judgement – what we accept as meaningful or valid – do not always rely on empirical experience or on the practical success of certain actions or beliefs. He questions the idea that experience necessarily underpins or legitimates the frameworks through which we judge and understand the world. Judgement, for Wittgenstein, is a form of practice, deeply embedded in human activities and forms of life, rather than being reducible to direct experience or its utility. In the next remark 132, Wittgenstein provides examples to illustrate this point further. He observes: "Men have judged that a king can make rain; we say this contradicts all experience. Today they judge that aeroplanes and the radio etc. are means for the closer contact of peoples and the spread of culture" (Wittgenstein, 1969, p. 19e). This comparison highlights the historical and cultural variability of human judgements. At one time, people believed that kings possessed the power to influence the weather — an assertion that, from the perspective of modern science, *contradicts all experience*. Nevertheless, such a belief was, for those who held it, a legitimate judgement within their cultural and conceptual framework. By contrast, Wittgenstein observes that contemporary society judges technological advancements, such as aeroplanes and radio, to be instruments for fostering closer human connections and the spread of culture. The success or widespread acceptance of these judgements does not make them immune to shifts in perspective over time. Here Wittgenstein emphasises that judgements are not simply grounded in experience, rather, they emerge from and are shaped by specific contexts, practices, and conceptual frameworks, i.e., shared *forms of life*.

In his 1931 diaries, Wittgenstein wrote that an "ethical proposition is a personal act" (Wittgenstein, 2003, p. 85). We are taught to regard something as good because others have said it is good. If it impresses a person and evokes admiration, the ethical proposition becomes meaningful to them; otherwise, it does not.

> An ethical proposition states "You shall do this!" or "That is good!" but not "These people say that this is good." But an ethical proposition is a personal act. Not a statement of fact. Like an exclamation of admiration. Just consider that the justification of an "ethical proposition" merely attempts to refer the proposition back to others that make an impression on you. If in the end you don't have disgust for this & admiration for that, then there is no justification worthy of that name. (Wittgenstein, 2003, p. 85)

Wittgenstein highlights that ethical propositions do not derive their force from factual justification but from their capacity to evoke a profound, personal response. Attempts to justify an ethical claim often trace back to the impression made by others or by particular values that resonate deeply with the individual. However, justification ultimately depends on one's own reactions – specifically, feelings of admiration or disgust. Without such personal responses, ethical propositions lack the weight or authority that might make them meaningful. In this way, Wittgenstein emphasises the *subjective* foundation of ethics, the significance of an ethical proposition is rooted not in external facts but in individual emotional engagement.

Almost the same can be said about the varieties of religious experience, echoing the title of W. James's book that had a significant influence on Wittgenstein. O'C. Drury recollected Wittgenstein saying: "The ways in which people have had to express their religious beliefs differ enormously.

All genuine expressions of religion are wonderful, even those of the most savage people" (Drury, 1996, p. 93). In the *Remarks on Frazer's Golden Bough*, Wittgenstein continued his reflections on this subject.

Was Augustine in error, then, when he called upon God on every page of the *Confessions*?

> But — one might say — if he was not in error, surely the Buddhist holy man was — or anyone else — whose religion gives expression to completely different views. But *none* of them was in error, except when he set forth a theory. (Wittgenstein, 1993, p. 119)

Wittgenstein also drew our attention to the variety of meanings attached to the concept of gods by different peoples. "Identifying one's own gods with the gods of other peoples. One convinces oneself that the names have the same meaning" (Wittgenstein, 1993, p. 119). However, this is entirely erroneous. In a discussion with Schlick on the concept of "value", Wittgenstein emphasised: "I would reply that whatever I was told, I would reject, and that not because the explanation was false but because it was an *explanation*" (Wittgenstein, 1979, p. 116); and further he continued: "What is ethical cannot be taught" (Wittgenstein, 1979, p. 117). According to O'C. Drury (1996, p. xii), for Wittgenstein, "The whole modern conception of the world is founded on the illusion that the so-called laws of nature are the explanation of natural phenomena." O'C. Drury argues that "philosophical clarity then arises when we see that behind every scientific construction there lies the inexplicable" (Drury, 1996, p. xii). In 1930, during a conversation with members of the Vienna Circle, Friedrich Waismann asked Wittgenstein the question: "Is the existence of the world connected with what is ethical?" Wittgenstein replied, "Men have felt that there is a connection and they have expressed it thus: God the Father created the world, the Son of God (or the Word that comes from God) is that which is ethical. That the Godhead is thought of as divided and, again, as one being indicates that there is a connection here" (Wittgenstein, 2003, p. 223, margin note). In *Philosophical Remarks* §54 (Wittgenstein, 1975, p. 84), Wittgenstein again pointed out, "What belongs to the essence of the world cannot be expressed by language." Language is limited; it "can only say those things that we can also imagine otherwise" (Wittgenstein, 1975, p. 84). He continued, "For what belongs to the essence of the world simply *cannot* be said. And philosophy, if it were to say anything, would have to describe the essence of the world" (Wittgenstein, 1975, p. 85). Yet, "the essence of language is a picture of the essence of the world"; and philosophy can "grasp the essence of the world, only not in the propositions of language, but in rules for this language which exclude nonsensical combinations of signs." On 10 October 1929 Wittgenstein wrote in code,[4] "What is good is also divine. Queer as it sounds, that sums up my ethics. Only something supernatural can express the Supernatural" (Wittgenstein, 1980, p. 3e). And four days later he left a remark in a normal script, "You cannot lead people to what is good; you can only lead them to some place or other. The good is outside the space of facts" (Wittgenstein, 1980, p. 3e). Moreover, I would like to apply to the words of Bouwsma, who concluded that the later Wittgenstein has not renounced his most important early "mystical" ideas: (1) "The sense of the world must lie outside the world"; (2) "In it there is no value, it must lie outside all happening and being-so. It must lie outside the world"; (3) "Ethics and aesthetics are the one"; (4) "The solution of the riddle of life in space and time lies outside space and time" (Bouwsma, 1986, p. 68).

Criticism of Dogmatism

Within the framework of pluralism concerning ethical, cultural, and religious systems, Wittgenstein regarded dogma as representing religion in a "bad sense". It demands, "Do this! – Think like that! – but it cannot justify this, and once it tries to, it becomes repellent" (Wittgenstein 1980, p. 29e). In *Philosophical Investigations* §373, Wittgenstein stated clearly about theology: it is a grammar (with its special rules) of its subject matter, religion. Theology thus becomes not a series of factual claims, but "grammatical remarks expressing rules for the use of theological terms in everyday religious discourse" (see Arrington, 2001, p. 172).

According to O'C. Drury's recollections (Drury, 1996, p. 86), Wittgenstein told him: "The symbolism of Christianity is wonderful beyond words, but when people try to make a philosophical system out of it I find it disgusting". Continuing his critique of dogmatic Christianity, Wittgenstein touched upon one of the most controversial aspects, namely, Christianity's offering of a historical narrative not based on historical truth. "It says, now believe!" (Wittgenstein, 1980, p. 32e). "But not, believe this narrative with the belief appropriate to a historical narrative, rather, believe, through thick and thin, which you can do only as a result of a life." Wittgenstein urged not to "take the same attitude to it as you take to other historical narratives! Make a quite different place in your life for it." He clarified that "historical proof (the historical proof-game) is irrelevant to belief".

4 See also MS 107, 192 and *Culture and Value*, rev. 2nd edition (Oxford, Basil Blackwell), 1998, page 5e.

Wittgenstein's view was that history and religion occupy different epistemological planes, corresponding to different spheres of knowledge (different language-games with their special grammar). We cannot apply historical methods to religion (unless studying the history of religion), nor can we use religious means, such as "religious belief", to interpret history. This resonates with Ernst Cassirer's concept of *symbolic forms*, in which Cassirer distinguished six basic forms of representation: language, myth, art, religion, history, and science. He argued that these categories form the consciousness of objects in the theoretical, intellectual sphere and are fundamental to shaping a worldview from the chaos of impressions. "Myth and art, language and science, are in this sense configurations towards being: they are not simple copies of an existing reality but represent the main directions of the spiritual movement, of the ideal process by which reality is constituted for us as one and many" (see Cassirer, 1955, *PSF*, V. II, p. 29; *PSF*, V. I, p. 107).

Wittgenstein refers to religious belief, using the example of the Gospels, as an "a*cceptance-as-true*." He clarifies that "we have quite different attitudes even to different species of what we call fiction" (Wittgenstein, 1980, p. 32e). In the notes published in *Philosophical Investigations*, Wittgenstein elaborates on his relation to religious questions, emphasising that they are a matter for people who have chosen this "*way of life*". For such individuals, these questions are not mere abstract "chatter", but are instead central to their lives. A religious question is either a question of life, or it is (empty) chatter. This language-game, one might argue, is played only with questions of life, much as the word "ouch" has meaning only as a cry of pain.

> I wish to assert that if eternal bliss holds no significance for my life or way of life, I need not trouble myself over it. If I am to rightly contemplate it, then my thoughts must bear a precise relation to my life; otherwise, my thoughts are either meaningless or my life is in danger. An authority that is ineffective, that I need not heed, is no authority. To rightly speak of an authority, I must also be dependent upon it. (Wittgenstein 2003, pp. 211–3)

In another conversation with O'C. Drury, Wittgenstein spoke extensively about his attitude to religion, particularly his view of Christianity in the form of Catholicism. He reiterated his idea concerning norms and customs, noting that true faith, for him, always transcended Church ceremonies. It was, rather, a personal relationship between an individual and God: "Make sure that your religion is a matter between you and God only" (Drury, 1984, p. 102). He reminded Drury that "Christianity is not a matter of saying a lot of prayers; in fact, we are told not to do that." Wittgenstein continued, expressing his thoughts more fully: "If you and I are to live religious lives, it must not be that we talk a lot about religion, but that our manner of life is different. It is my belief that only if you try to be helpful to other people will you, in the end, find your way to God" (O'C. Drury, 1984, p. 114). At the end of 1931, Wittgenstein wrote in his diaries: "Christianity is really saying: let go of all intelligence" (Wittgenstein, 2003, p. 139). In 1937, he added: "To get rid of the torments of the mind, that is to get rid of religion" (Wittgenstein, 2003, p. 199). Wittgenstein explained to Drury the essence of the normative conflict that made it impossible for him to be a Roman Catholic:

> It is a dogma of the Roman Church that the existence of God can be proved by natural reason. Now, this dogma would make it impossible for me to be a Roman Catholic. If I thought of God as another being like myself, outside myself, only infinitely more powerful, then I would regard it as my duty to defy him. (Drury 1996, pp. 107–8)

Wittgenstein continued, asserting that "religion takes many forms; there are similarities, but there is nothing common among all religions" (Bouwsma, 1986, p. 55). He remarked that he never objected to anyone's religious beliefs, stating, "Believe whatever you can" (Bouwsma, 1986, p. 56).

Conclusion

This paper has examined Wittgenstein's critique of the dispositional theory of values, and argues that later Wittgenstein was not "full-body conventionalist", showing his nuanced and sometimes conflicting understanding of aesthetics, ethics, ethical systems, and world-views. Wittgenstein rejected the idea that aesthetics could be reduced to personal taste. By distinguishing between taste and aesthetic judgement, he emphasised that while tastes and conventions fall within the domain of psychology, aesthetics is concerned with "rules of harmony" and ideals that transcend mere subjective preference. He argued that ethical, aesthetic, and religious expressions frequently involve a characteristic "misuse of our language", in which terms acquire meanings that, while similar to, are distinct from their ordinary usage.

Moreover, Wittgenstein distinguished "ethics" which is similar to aesthetics and transcend conventions from ethical systems as descriptions of societal morals and customs, arguing that ethical propositions cannot be equated with empirical statements or sociological observations. The use of rules is determined by their grammar, and ethical systems, like language-games, embody a plurality of culturally embedded frameworks. While ethical judgements involve the adoption of specific criteria, no single system can claim universal validity. As Wittgenstein observed, when someone asserts that something is good, they are making a judgement of value. This plurality of ethical systems, each with its own grammar, precludes the possibility of declaring one system inherently superior to another. But returning to "ethics" in the first sense as the representation of higher values, Wittgenstein in his later writings again repeats his early ideas that "The good is outside the space of facts" (Wittgenstein, 1980, p. 3e).

Wittgenstein extended his pluralistic perspective on the level of language-games and societal norms and customs to world-views and variety of religious experience, asserting that all views are equally significant to those who hold them. For Wittgenstein, ethical judgements are deeply personal acts, shaped by cultural learning and individual experience, with no universal or teachable doctrines. Moreover, he recognised a plurality of ethical systems and world-views, rejecting the notion of a single superior system. This pluralistic view underscores his broader understanding of ethics and aesthetics as grounded in human practices yet *transcending* mere subjective or cultural conventions. "You cannot lead people to what is good; you can only lead them to some place or other" (Wittgenstein, 1980, p. 3e). What is good, for Wittgenstein, cannot be learned, for it is divine, and thus supernatural. "What is good is also divine. Queer as it sounds, that sums up my ethics. Only something supernatural can express the Supernatural" (Wittgenstein, 1980, p. 3e). It lies outside the realm of facts, conventions, attitudes, and tastes, and is inexpressible – i.e., mystical.

References

Amesbury, R. (2020). *Ethics after Wittgenstein*. London/New York: Bloomsbury Academic.

Anscombe, G. E. M. (1958). *Intention*. Oxford: Basil Blackwell.

Austin, J. L. (1950). "Truth." Proceedings of the Aristotelian Society, Supplementary Volumes, 24, 111–128.

Ben-Menahem, Y. (2006). *Conventionalism: From Poincare to Quine*. Cambridge University Press.

Berg, Á. 2024. Was Wittgenstein a radical conventionalist? *Synthese* 203 (2), 1–31.

Bouwsma, O. K. (1986). *Wittgenstein, Conversations, 1949–1951*. Indianapolis, USA: Hackett.

Cassirer, E. (1955) *The Philosophy of Symbolic Forms, Volume 1: Language*. New Haven: Yale University Press.

Cassirer, E. (1955) *The Philosophy of Symbolic Forms, Volume 2: Mythical Thought*. New Haven: Yale University Press.

Christensen, A.-M. S. (2007). "Ethical Objectivity and Wittgenstein." *Philosophical Investigations*, 30/3, 219–236.

Citron, G. (ed.) (2015). A Discussion Between Wittgenstein and Moore on Certainty (1939), From the Notes of Norman Malcolm. *Mind,* 124(493), 73–84.

Drury, M. O'C. (1984). Conversations with Wittgenstein. In R. Rhees (ed.), *Recollections of Wittgenstein* (pp. 97–171). New York: Oxford University Press.

Drury, M. O'C. (1996). *The Danger of Words and Writings on Wittgenstein*. Bristol: Thoemmes Press.

Dummett, M. (1978). *Truth and Other Enigmas*. London: Duckworth.

Dummett, M. (1959). Wittgenstein's philosophy of mathematics. *The Philosophical Review,* 68(3), 324–348.

Glock, H.-J. (2003). The linguistic doctrine revisited. *Grazer Philosophische Studien,* 66(1), 143–170.

Glock, H.-J. (2008). Necessity and language. In defence of conventionalism. *Philosophical Investigations,* 31(1), 24–47.

Hanna, Robert. 2008. *Kant in the Twentieth Century*, in: D. Moran (Ed.), *The Routledge Companion to Twentieth Century Philosophy*. London/New York: Routledge, 146–178.

Hare, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.

Hämäläinen, Nora. 2015. *Descriptive Ethics: What It Is and Why We Need It*. Cham: Palgrave Macmillan.

Iczkovits, Y. (2012). *Wittgenstein's Ethical Thought*. Basingstoke: Palgrave Macmillan.

Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press.

McDowell, J. (1984). *Wittgenstein on Meaning*. Oxford: Clarendon Press.

O'Connor, P. 2000. *Oppression and Responsibility: A Wittgensteinian Approach to Social Practices and Moral Theory*. University Park: Pennsylvania State University Press.

Putnam, H. (1979). Analyticity and apriority: Beyond Wittgenstein and Quine. *Midwest Studies in Philosophy,* 1(4), 423–441.

Quine, W. van O. (1966). Truth by convention. In: W. van O. Quine (Ed.), *The Ways of Paradox and other Essays* (pp. 70–99). New York: Random House.

Raatzsch, Richard. 2010. *Apologizing Evil: Iago's Case*. Princeton: Princeton University Press.

Rhees, R. (1984). *Recollections of Wittgenstein*. Oxford: Oxford University Press.

Topey, B. (2019). Correction to: Linguistic convention and worldly fact: Prospects for a naturalist theory of the a priori. *Philosophical Studies* 176 (7), 1753–1755.

Topey, B. (2019). "Linguistic convention and worldly fact: Prospects for a naturalist theory of the a priori." *Philosophical Studies* 176 (7), 1725–1752.

Waismann, F. (1979). *Wittgenstein and the Vienna Circle*. New York: Barnes & Noble.

Warren, J. (2015). The possibility of truth by convention. *The Philosophical Quarterly,* 65(258), 84–93.

Winch, P. (1958). *The Idea of a Social Science and Its Relation to Philosophy*. London: Routledge & Kegan Paul.

Wisnewski, J. J. (2007). *Wittgenstein and Ethical Inquiry: A Defense of Ethics as Inquiry*. London: Continuum.

Wittgenstein, L. (1993). A Lecture on Ethics. In J.C. Klagge and A. Nordman (eds.) *Philosophical Occasions, 1912–1951* (pp. 37–44). J. Indianapolis: Hackett Publishing.

Wittgenstein, L. (1980). *Culture and Value*. Oxford: Basil Blackwell.

Wittgenstein, L. (1967). *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*, [1938–1946]. Berkeley and Los Angeles: University of California Press.

Wittgenstein, L. (1969). *On Certainty*. New York: Harper Torchbooks.

Wittgenstein, L. (1974). *Philosophical Grammar*. Oxford: Blackwell.

Wittgenstein, L. (1993). *Philosophical Occasions*: *1912–1951*. Indianapolis and Cambridge: Hackett.

Wittgenstein, L. (1975). *Philosophical Remarks*. New York: Barnes and Noble.

Wittgenstein, L. (2003). *Public and Private Occasions*. Md: Rowman & Littlefield.

Wittgenstein, L. (2005). *The Big Typescript*. Oxford: Blackwell Publishing.

Wittgenstein, L. (1998). *The Blue and Brown Books*. Oxford: Blackwell.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Kegan Paul.

Wittgenstein, L. (2016). *Wittgenstein: Lectures, Cambridge 1930–1933: From the Notes of G. E. Moore*. Cambridge: Cambridge University Press.

Wittgenstein, L. (2001). *Wittgenstein's Lectures, Cambridge 1932–1935: From the Notes of A. Ambrose and M. MacDonald.* New York: Prometheus Books.

Wittgenstein, L. (1968). Notes for lectures on private experience and sense data. *Philosophical Review* 77 (3): 275–320.

Wittgenstein, L. (1970). *Zettel*. Berkeley: University of California Press.

Wright, C. (1980). *The Debate over the Foundations of Wittgenstein's Philosophy*. Philosophical Quarterly, 30(119), 193–213.

CIP – Каталогизација у публикацији –
Народна библиотека Србије, Београд

1(082)   7.01(082)   159.9(082)   17(082)

BALKAN Analytic Forum (2 ; 2025 ; Beograd)

Dispositions & dispositions and values : conference proceedings / Balkan Analytic Forum, 10–20. X 2024. Belgrade, Serbia ; edited by Miroslava Trajkovski, Emily C. McWilliams. – Belgrade : University, Faculty of Philosophy : Center for Contemporary Philosophy, Balkan Analytic Forum, 2025 (Belgrade : Službeni glasnik). – 299 str. : ilustr. ; 24 cm

"Welcome to the second volume of the Balkan Analytic Forum's proceedings, featuring contributions from BAF2: Dispositions and BAF2+: Dispositions and Values." --> Preface. – Tiraž 200. – Str. 9: Preface / Emily C. McWilliams. – Napomene i bibliografske reference uz radove. – Bibliografija uz svaki rad.

ISBN 978-86-6427-381-7 (UFP)

а) Филозофија – Зборници
б) Филозофија уметности – Зборници
в) Вредности – Зборници

COBISS.SR-ID 175158537

The Balkan Analytic Forum aims to bring together experts in analytical philosophy from the Balkans to exchange their ideas, but it is also open to approaches that establish connections between analytical philosophy and other philosophical traditions, as well as to interested experts from other parts of the world. The activity of the forum is to establish, through conferences and accompanying publications, a platform for discussion where scientists from the Balkans, and all those interested in analytical philosophy, can meet regularly and present the texts they are currently working on and their new publications. The mission is to carry out basic, applied and developmental research in the domain of analytical philosophy; to publish the results of this scientific research and professional work; to include young researchers and doctoral students at the Faculty in the implementation in this through participation in the programs implemented by the Center for contemporary philosophy – Balkan Analytic Forum of the University of Belgrade, Faculty of Philosophy (https://www.f.bg.ac.rs/balkanski_analiticki_forum); to participate in the organization of gatherings, symposiums, professional meetings and workshops for the purpose of training researchers in the field of analytical philosophy; as well as to cooperate with other institutions in the country and abroad, especially with countries from the Balkans.

Miroslava Trajkovski

The concept of "disposition" has been central to the social sciences for the past fifty years, with applications ranging from Bourdieu's "habitus" to Foucault's "dispositif". However, analytic philosophy has shown comparatively less interest—until recently. That may be starting to change. This highly recommended volume explores the philosophical concept of "disposition" as it applies to areas including logic, ethics, and aesthetics.

Michał Kozłowski, *University of Warsaw, Poland*

This volume collects papers from the *Balkan Analytic Forum* (BAF) conference held in Belgrade in 2024, which focused on the role of dispositions and values in different areas of philosophy and science. The contributions move beyond the narrow frameworks in which these issues are often discussed, offering new perspectives and approaches. They are intended to be both insightful and illuminating for a broad range of phenomena, as well as to encourage further enquiry into these topics.

Duško Prelević, *University of Belgrade, Serbia*

This second volume of the *Balkan Analytic Forum* brings together leading and emerging voices in contemporary philosophy to explore the concept of dispositions in its many dimensions. Essays address evidence, ontology, virtue, and emotion, offering fresh insights into how dispositions shape thought, character, and culture. Bringing together international and regional voices, the collection reflects the Balkan Analytic Forum's mission to foster rigorous dialogue across traditions and disciplines.

Radmila Jovanović Kozłowski, *University of Belgrade, Serbia*

ISBN 978-86-6427-381-7

9 788664 273817

UNIVERSITY OF BELGRADE
FACULTY OF PHILOSOPHY