Bryce Huebner
Department of Philosophy
Georgetown University

Original Scientific Paper
UDK 165.194:159.955

PICTURING, SIGNIFYING, AND ATTENDING[1]

Abstract. *In this paper, I develop an empirically-driven approach to the relationship between conceptual and non-conceptual representations. I begin by clarifying Wilfrid Sellars's distinction between a non-conceptual capacity to picture significant aspects of our world, and a capacity to stabilize semantic content in the form of conceptual representations that signify those aspects of the world that are relevant to our shared practices. I argue that this distinction helps to clarify the reason why cognition must be understood as embodied and situated. Drawing on recent models of attention and valuation, I then argue that the human brain constructs a dynamic model of the world that it has encountered, encoding higher-level regularities in the form of linguistically structured representations. And I conclude by arguing that this approach to cognition provides a set of critical resources for understanding the situated nature of social cognition.*

Key words: *Situated Cognition; Sellars; Cybernetics; Attention; Social Cognition*

Introduction

Nutmeg is becoming impatient. It is early in the afternoon, and she is starting to feel hungry. She meows and paws at me. So I try to remind her that the auto-feeder will deliver her food at precisely 15:00; but she doesn't seem to listen or care. As you might have guessed, Nutmeg is a cat; so she isn't a language user, and it's unlikely that her thoughts rely on symbolic states with precise conceptual and propositional structure. She seems to have a

1 The ideas in this paper have been percolating for a very long time. They were first presented at a conference on *Sellars's Legacy,* at the American University of Beirut (26 May 2015). Various parts of this project were then developed in papers that I presented at The University of Maryland, Associazione Italiana di Scienze Cognitive, The George Washington University, Virginia Tech, The NEH Summer Institute on Presupposition and Perception, Minds Online, Universität Stuttgart, Université du Québec à Montréal, The University of Michigan, Georgia State University, and Mind & Life Europe. I am thankful for the discussions I had in each of these places, and for the chance to learn from many amazing philosophers and cognitive scientists. I would also like to offer a special thanks to Rebecca Todd for helping me to understand the importance of implicitly guided attention; I would like to thank Carl Sachs for numerous conversations about Sellars—and many other issues; and I would like to thank Carl Sachs, Ben Elzinga, Dan Williams, and two anonymous referees for this journal for their helpful comments on earlier drafts of this paper. Finally, and most importantly, I would like to thank Ruth Kramer for helping me work through the toughest parts of this paper, and for always being there when I struggle to find a way forward.

 doi: 10.5937/BPA1831007H

fairly good sense of where her food will be, and roughly when it will arrive; she also seems to have numerous learned expectations about the micro-world in which she lives; and there are many things that really do seem to matter to her—including access to food and human attention. But so far as I can tell, she doesn't have anything like a mental language that consists of "word-sized concepts, sentence-sized intentional states and argument-sized inferences" (Williams 2018, 153). The neural representations that have been observed in the brains of nonhuman animals typically take the form of topographic maps, representations of local motion, efference copies, and forward models (Thomson & Piccinini 2018). And while such representations are sufficient to guide goal-directed behavior in specific contexts, they offer little insight into the human ability to "impose stability, order, and uniformity upon a conception of the world" (Akins 1996, 368).

While there are notorious difficulties with appeals to conceptual representations (Akins 1996, 367ff; Chomsky 1995; Thompson 2010), psychological explanations of human behavior commonly appeal to symbolic states that are "tailor-made for semantic interpretation" (Egan 2019, 247). This shouldn't be surprising, as such states can be organized to yield networks of computational processes that can "directly encode and exploit the kinds of information that a human agent might consciously access when trying to solve a problem" (Clark 2014, 35; Newell & Simon 1956, 1976). And they seem to provide a nice bridge between our folk-ontology and a computational model of the mind (Schneider 2009; Salisbury & Schneider 2019). Of course, most cognitive scientists and many philosophers acknowledge that appeals to conceptual states are an idealization. But the assumption that human thought relies upon word-shaped concepts and sentence-shaped thoughts persists in many domains, shaping psychological hypotheses, and generating intractable disputes over the nature of mental representation. So we seem to face a dilemma: we can either treat the brain as a computational system, which operates over identifiable neural representations that have little in common with the folk understanding of thought; or we can work to preserve the model of thought that is central to folk-psychology, while abandoning hope when it comes to the computational theory of mind.

There are long running debates about these issues. And I don't intend to address them head on. But I hope to make headway on these issues by providing an empirically-driven approach to the relationship between conceptual and non-conceptual representations; some aspects of this framework will be familiar, but by highlighting the importance of attention and valuation, I show that there is a way of seeing the brain as a dynamic system, which is shaped by our ongoing interactions with the world, and which also encodes higher-level regularities that take the form of linguistically structured representations. In making this case, I begin with the model of cognition that was developed by Wilfrid Sellars (1960). Sellars distinguished two kinds of

cognitive capacities: a non-conceptual capacity to *picture* significant aspects of our world; and a capacity to stabilize semantic content in the form of conceptual representations that *signify* those aspects of the world that are relevant to our shared practices. While this may seem like an odd place to start, I contend that the approach that Sellars advances helps to clarify the reason why cognition must be understood as embodied and situated, even if we retain a relatively traditional approach to cognitive processing. From here, I move to the implicit forms of valuation and attention that organize patterns of thought and behavior non-conceptually; and I argue that this addition to Sellars model can help to provide a more contemporary foundation for understanding the nature of conceptual thought. And I conclude by showing how this approach to cognition can be applied to help clarify the situated nature of social cognition.

1. Picturing the world

In one of the earliest philosophical articulations of the computational theory of mind, Sellars (1960) describes a robot that encodes information about the world in memory, by 'printing sentences on its tape'. This robot wanders "around the world, scanning its environment, recording its 'observations', enriching its tape with deductive and inductive 'inferences' from its 'observations' and guiding its 'conduct' by 'practical syllogisms' which apply its wired in 'resolutions' to the circumstances in which it 'finds itself'" (Sellars 1960, §39). Over time, the robot develops a better 'understanding' of the world, and becomes more attuned to the aspects of the world that are relevant to its needs. This robot looks a lot like the kind of learning machine that Alan Turing (1950) posits in the final section of "Computing machinery and intelligence". And at least initially, it seems to rely on symbolic forms of representation, which can be nicely organized into a language of thought. But as the generous use of scare quotes suggests, Sellars was skeptical of this characterization of the robot as a conceptual system; and his distinction between picturing and signifying is intended to provide a way of decoupling the computational features of thought from the capacity for conceptual representation (Sellars 1960, §32).[2] The significance of this claim will become clearer over the course of this paper; but for now, the important thing to note is that Sellars's distinction turns on an account of the embodied strategies that cognitive systems develop as they learn to engage with aspects of the world that matter to them. Sellars argues that cognitive systems develop models of the world that they inhabit using simple forms of error-driven learning; the resulting models are highly structured, but they are not conceptually organized—they are holistically structured models that are constantly

2 For alternative discussions of picturing, which explore Sellars's claims in more detail, see Levine (2007), O'Shea (2007), and Sachs (2018).

updated in light of new information. That said, they provide the foundation for conceptual thought. But to see how they do so, it is necessary to first clarify the dynamic structure of these models.

1.1 What is a picture?

Throughout his career, Sellars (1956; 1960; 1974; 1981) explored a variety of different ways of understanding capacities for learning and self-regulation. He was an avid reader in the cognitive and behavioral sciences, and he often drew on something like Edward Tolman's *purposive behaviorism* (cf., Olen 2018). Tolman (1932) had argued that learning is an active and goal-directed process: animals explore different strategies for pursuing things that matter to them (e.g., finding food and cuddling with friends); they use their successes and failures to improve their understanding of the world; and over time, they construct cognitive maps that will effectively guide their behavior, at least in familiar contexts (Tolman 1948). Likewise, Sellars (1981 §56) argues that "to be a representational state, a state of an organism must be the manifestation of a system of dispositions and propensities by virtue of which the organism constructs maps of itself in its environment, and locates itself and its behavior on the map"; he claims that cognitive maps play a critical role in the guidance of flexible and adaptive forms of behavior; and, he appeals to forms of feedback-driven learning to explain how we construct "an increasingly adequate and detailed picture of" the world (Sellars 1960, §40). But Tolman never explained how cognitive maps were realized in the brain; and at points, he seems to treat them as internal images that show up to the animal who uses them. By contrast, Sellars argues that cybernetic theory throws "light on the way that cerebral patterns and dispositions picture the world" (Sellars 1960, §59). And this is where his approach diverges from standard forms of machine functionalism, yielding an embodied and situated understanding of thought and agency. Put much too simply for now, a cybernetic approach to cognition draws our attention to the control of purposive behavior; it highlights the importance of ongoing feedback in the production of resilient dynamic relations between an organism and the world; and while cybernetic systems can be designed to accommodate symbolic representations and person-level inferences, they tend to be more concerned with control over the actions that allow a system to survive and flourish in its natural environment.

While Sellars doesn't go into detail on any of these points, they do play a prominent role in his discussion of embodiment (as I argue in the next subsection). And they do seem to be implicit in the accounts of cognition that he draws upon. Specifically, his reference to cybernetic theory as it would have been understood in the 1960 suggests a view of picturing that depends on informational relations, which can be implemented in the network structure of a brain. In the early 1940s, Warren McCulloch & Walter Pitts (1943) argued that the all-or-nothing character of neural activity allowed individual neurons

to function as logic gates, which could be cyclically organized to represent logically structured propositions. Skeptics quickly noted that neural activity could not be captured by digital flows of propositional states (Abraham 2019); and a variety of cybernetic alternatives rapidly emerged, focusing on the nature of behavioral control in animals and machines (Wiener 1948). These approaches retained the commitment to mechanisms that detected, processed, interpreted, and stored information at multiple points in the brain. For example, Donald Hebb (1949) argued that associative learning could be implemented in a system that relied on the strengthening of connections between neurons that were active in concert, inducing lasting cellular changes that could facilitate the storage of information. Frank Rosenblatt (1958) used feedback-driven learning to show how classificatory information could be acquired and stored in the dynamic connections within a neural network. And having learned that cells are responsive to specific features of objects (e.g., length, orientation, contrast), Oliver Selfridge (1958) developed a pandemonium architecture, which used multiple 'demons', working in parallel to extract meaningful patterns from noisy signals.

As Sellars (1981, §64) notes, however, the storage and processing of information is only part of the story when it comes to the nature of mental representation: the cognitive maps we construct and use exploit a network of interconnected states, which drive motor activity, and generate reliable strategies for getting around in the world. And by the late 1950s, a similar insight had inspired research into the frog's capacity to reliably detect, track, and consume fast moving insects. Building on the insights that had led Selfridge to develop the Pandemonium architecture, Jerome Lettvin and his colleagues (1959) argued that a plausible understanding of mental representations should appeal to the role of interacting systems in the guidance of goal-directed behavior. But they moved beyond computational models, to show how such operations could be implemented in a biological brain. They identified cells that were responsive to small dark shapes moving across the visual field; they suggested that computations carried out by these cells could be specified in terms of operations over edges and contrasts, curvature, movement, and luminance; and they argued that information from these cells was stored in a retinotopically organized map in the superior colliculus, which facilitated control over food-seeking behavior.

The representations posited by this model were deeply tied to adaptive behavior, and they were characterized in terms of feature detectors, retinotopic maps, and mathematical functions. Moreover, the systems that Lettvin and his colleagues posited were tightly coupled to the flow of information through specific neural systems, yielding a highly promising approach to linking neuroscience and behavior (Maturana et al 1959; Lettvin et al 1959). But just as importantly, these approaches described the relevant class of computations in mechanical terms; and any heuristic gloss of what the frog was representing (e.g., flies, or fast moving insects) would necessarily

go beyond what the data could support (cf., Egan 2014). Consequently, while these models revealed important facts about how frogs represent the world, they didn't reveal anything about their ability to track the category 'insect' (this is the primary sense in which they are non-conceptual systems). Indeed, the mechanisms that were discovered in their eyes and brains only seemed to track differences in motion and light. Moreover, these models showed that the ability to reliably track these kinds of differences could be explained without attributing any sensitivity to any conceptual categories to the frog itself.

1.2 Recordings and embodiment

Sellars never addresses the frog's capacity to picture the world. But even if he was unaware of this research, he was concerned with the kind of informational isomorphism that we find in the states of a neural network, and the mechanisms in the superior colliculus that preserve features of the world that are most salient to behavioral guidance. To see why, consider an analogy that he offers to the way that the groove on a vinyl record pictures a musical performance (Sellars 1960, §40). The groove is caused by the acoustic properties of a specific performance; and it records the sonic profile of the performance, using a function that maps acoustic differences onto differences in groove depth; and this recording can be recovered, using an inverting function to map differences in groove depth onto a sonic output. But while the record is always machine-readable, it only becomes person-usable when it is placed on a turntable, with the right kind of needle, rotating at the right speed, and sending the right kind of signal through an amplifier to a set of speakers. As I read him, this is why Sellars (1960, §40) suggests that this picture "cannot be abstracted from the procedures involved in making and playing the record".[3]

Building on this analogy, we can see cognitive maps as recordings, which are produced by a mechanism that maps perceptible differences in evaluative salience onto differences in the structure of a neural network. In picturing the world, a cognitive system takes in perceptual information, represents it using an isomorphic relation, and uses this representation to guide its purposive

3 Compare Frances Egan's (2014, 116) description of the computation of the addition function: "A physical system computes the addition function just in case there exists a mapping from physical state types to numbers, such that physical state types related by a causal state-transition relation (p1, p2, p3) are mapped to numbers n, m, and n+m related as addends and sums. Whenever the system goes into the physical state specified under the mapping as n, and then goes into the physical state specified under the mapping as m, it is caused to go into the physical state specified under the mapping as n+m." Importantly, this way of characterizing computation requires nothing more, and nothing less than: 1) a functional mechanism that can process variables that change states (e.g. patterns of neural spiking), 2) in accordance with rules that map inputs to outputs, 3) in ways that are sensitive to the properties of these variables, and to differences between different portions of them (Piccinini 2015; Piccinini & Bahar 2013; Piccinini & Scarantino 2011).

behavior. But like the grooves on a vinyl record, the patterns in a neural network do not picture something just by being present in the structure of a system; picturing "involves the manner in which the patterns...are added to, scanned, and responded to by the other components" of the system (Sellars 1960, §40). And the isomorphic relations that encode information about the world only become a way of picturing things "by virtue of the physical habitus of the" system, that is "by virtue of mechanical and electronic propensities which are rooted, ultimately, in its wiring diagram" (Sellars 1960, §40).

This may seem like a strange way to phrase this claim. After all, the term 'habitus' is typically associated with the sociological research of Pierre Bourdieu. And it is commonly used to identify the network of embodied dispositions that organize an individual's perception of, and actions within, the social world. But Sellars's use of the term derives from the work of Thomas Aquinas.[4] And here, an agent's *habitus* is understood as a mode of being, which arises through their intentional and purposeful use of an extrinsic thing, in a way that "actualizes one of the open-ended range of potentialities for engaging with the world engendered by human reason" (Spencer 2015, 121). For example, my use of the *rakweh* that I bought years ago in Beirut actualizes my ability to make coffee in a specific way; and it does so because my (learned) capacities for coffee-making fit with the function of the *rakweh*, in a way that actualizes a specific range of coffee-making activities. Extending these claims more broadly, we might say that picturing the world actualizes the ability to act in specific ways; and that it does so because we possess learned capacities that fit with the world, and that actualize specific forms of rationally structured action. To acquire a way of picturing the world is thus to become attuned to particular aspects of the world, and to orient our action toward them; but just as importantly, this structure of attunement is an embodied and situated strategy for engaging with the world, which constrains the sources of information that an agent will track and respond to, while delimiting their possibilities for acting.[5]

The mechanisms that produce such pictures must facilitate attunement to salient aspects of the world. They must produce internal states, which are isomorphic to salient features of the world; and in virtue of this fact, they must be able to guide purposive behavior. Finally, keeping with Sellars's appeal to cybernetics, we should see these mechanisms as operating by changing connection weights between neurons, altering their tonic or phasic firing rates, or stabilizing isomorphic relations between the states of a brain and state of the world. And this brings us to Sellars's core insight: picturing

4 "Being and being known" was originally published in the Proceedings of the American Catholic Philosophical Association, and Sellars claims that it is an exploration of "the profound truth contained in the Thomistic thesis that the senses in their way and the intellect in its way are informed by the natures of external objects and events".

5 For a far more compelling defense of a nearby perspective, see Kukla (2017).

is an *activity*, which depends on the use of a cognitive map to actualize one of an agent's capacities for acting in an embodied, embedded, and situated way; and since cognitive maps provide a representation of an agent's place in the world that *they have encountered,* they will always be deeply tied to that agent's capacities for action (cf., Sellars 1978 §28–29).[6] This insight also motivated much of the initial research in cybernetic theory. As Kenneth Craik (1943, 61) famously puts this point, if a cognitive system "carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilise the knowledge of past events in dealing with the present and future, and in every way react in a much fuller, safer, and more competent manner to the emergencies which face it". But at the same time, such models must be open-ended and revisable. They cannot rely on static symbolic representations, and they cannot be decoupled from the world in which they are embedded. And by focusing on these aspects of Sellars's model, we come to a more contemporary account of picturing.

2. A neo-Sellarsian approach to picturing

Thus far, I have argued that picturing is an activity that depends on the use of a cognitive map to actualize specific capacities for acting. In acquiring a way of picturing the world, an agent becomes attuned to salient phenomena, in ways that yield embodied and situated strategies for action, constrain the sources of information that an agent tracks and respond to, and delimit an agent's possibilities for acting. But at the same time, such capacities must be flexible, open-ended, and revisable. This is something that Sellars seems to have recognized, but in thinking about the nature of picturing as a foundation for nonconceptual thought, the salience of this flexibility is crucially important. My aim in this section is to show that recent research on evaluative learning and the implicit guidance of attention provide a way of thinking about how these capacities emerge and stabilize in human minds. And while the underlying processes aren't exactly maps or pictures, they do record information about the world in patterns of neural connectivity, and they sustain the kinds of behavioral capacities that I have been discussing. So they capture the core aspects of Sellars's theory of picturing. That said, my discussion of attention and valuation go beyond anything that Sellars ever said. And they draw us closer to redeeming these cognitive capacities in the coin on cybernetic and neuroscientific data.

---

6 Here, too, Sellars seems to be following Tolman (1948, 192), who argues that mind is "more like a map control room than it is like an old fashioned telephone exchange. The stimuli, which are allowed in, are not connected by just simple one-to-one switches to the outgoing response. Rather, the incoming impulses are usually worked over and elaborated in the central control room into a tentative, cognitive-like map of the environment. And it is this tentative map, indicating routes and paths and environmental relationships, which finally determines what responses, if any, the animal will finally release."

There is a broad consensus that typical humans rely upon a network of interacting systems—including the basal ganglia, amygdala, and anterior insula—to evaluate multiple kinds of information in parallel (cf., Gardner et al submitted; Gershman in press). Some of these systems track fluctuations in the value of rewards, others monitor changes in the probability of gains and losses, and others adjust subjective estimates of risk and uncertainty in light of experienced feedback (Montague et al 2012; Adolphs 2010). For example, the tonic and phasing spiking activity of dopaminergic neurons in the basal ganglia represent the predicted value and distribution of rewards; these patterns of activity are adjusted when rewards are better or worse than expected; and these adjustments continue until the value and distribution of rewards are predicted accurately (Schultz, 1998, 2010). Similar systems in the amygdala and striatum generate and revise predictions about aversive stimuli (Delgado et al 2008). And recent data suggest that subsecond dopamine fluctuations in the human striatum facilitate the computation of expected as well as merely possible outcomes (Kishida et al 2016). These systems sustain patterns of attunement to the evaluatively salient aspects of the world through Pavlovian and classical conditioning; their behavior is best understood computationally, but they do not rely on symbolic representations with determinate conceptual content. That said, they do play a critical role in orienting behavior toward evaluatively significant phenomena. And so long as the rate and value of risks, rewards, and threats remain relatively stable, these systems will tend to produce viable capacities to evaluate current and counterfactual situations (Railton 2014).

Like the cutting of grooves into a vinyl record, the activity of these computational systems shapes patterns of neural response, in ways that can be mapped onto properties of the world (e.g., reward, threat, and risk). While it would be a mistake to think that these systems yield anything like a geometric or relational map of the world, they do facilitate the storage of evaluatively salient information. They allow typical humans and many other animals to represent the rewards, threats, and risks they have previously encountered. And they orient patterns of thought and behavior, by constraining the sources of information that a typical human will track and respond to (for reviews in different domains, see Crockett 2013; Cushman 2013; Haas 2017; Huebner 2016). Importantly, these systems also play a critical role in the guidance of implicit attention (Anderson 2016; Todd and Manaligod 2017). And recent approaches to attention have claimed that they are integrated into a priority map of an attentional landscape, which is shaped by ongoing competition for selection between networks of value-driven, goal-directed, and stimulus-driven influences (Anderson 2016, 32; Awh et al 2012).

Many forms of attention are fast and automatic, and they appear to be shaped by lingering biases of selection history, especially by past experiences of reward (Theeuwes 2018). And according to one plausible hypothesis about how this occurs, dopaminergic signals enhance the perception of

evaluatively salient information in sensory cortex, as sensory states that are predictive of reward are prioritized across situations, sensory modalities, and stimulus properties—including locations, complex objects and their properties, and characteristics of experienced scenes. On this model, the resulting enhancement of cortical signaling then biases competition between sensory inputs, in ways that affect the salience of things in the world. This hypothesis gains support from the fact that signals from the reward system do seem to influence attention in ways that reflect "formal models of reward learning, including reflecting a common neural currency for value that scales with the amount of reward available in the current task" (Anderson 2016, 35). But these data are also consistent with the hypothesis that the reward system estimates the current state of the animal and its environment, and shapes the interpretation of sensory information in ways that produce patterns of perceptual orienting and behavioral response (Krauzlis et al 2014). According to this approach, attention is an effect of the reward system attempting to interpret sensory data, in light of prior knowledge, and the agent's current state. And intriguingly, connections between the superior colliculus and the basal ganglia provide a plausible point where information from sensory systems might be integrated with evaluative information. Finally, there is some reason to believe that a priority map is realized by a network of retinotopically organized structures in the parietal cortex, allowing an agent to "represent the task relevance, learned value, and physical salience of stimuli" in a single representational structure (Anderson 2016, 32).

Each of these perspectives suggests that strategies for processing current information will often reflect past experiences of reward, yielding habitual forms of attention that are anchored to learned patterns of evaluation (Anderson 2016, 35). However, we are always in some kind of affective and evaluative state, no matter what is happening, and no matter how neutral we feel (Lindquist 2013, 361). And this matters since the evaluative state we are in when we encounter a risk, a threat, or a reward will have a significant impact on how we interpret its salience, and in what information we encode for future action. Hungry people tend to be more attentive to the food that they see, and this enhances their food-related memories (Talmi et al 2013); people who are nervous about snakes tend to see them everywhere while hiking on a trail (Machery 2016); and the experience of economic precarity tends to enhance the salience of economic information (Shah et al 2018), and perhaps even shape the way that people perceive race (Krosch & Amodio 2014). These effects are important, but not just because of the shifts in what people perceive in the moment. To the extent that these tendencies shape what people notice, what they remember, and what they imagine to be possible, they will impact the structure of the picture that someone forms of the world (cf., De Brigard et al 2017, 2018). Put somewhat differently, such effects are part of the history that implicitly guides attention, and that determines how an agent will behave in any future context. And this will be true even if it turns out that perceptual systems have relatively stable, and relatively modular properties (Gross 2017; Machery 2016).

To explain these more pervasive kinds of effects, Rebecca Todd and her colleagues have developed a model of attention that posits priority maps, which are shaped by numerous aspects of an agent's 'history'. (Kryklyvyy & Todd 2018; Todd & Manaligod 2017) According to this model, statistically stable features of an agent's environment will tend to be prioritized, independently of their intrinsic salience or goal relevance (Zhao et al 2013); and this is because people are learning to map the structure of the world that they encounter, by adjusting their response to attentionally salient phenomena. Likewise, reward-driven learning helps to guide attention toward rewarding stimuli, and away from aversive stimuli; this helps an agent orient their behavior within a highly structured evaluative landscape. And in some cases, attentional landscapes can become attuned to the co-occurrence of related features or objects, in ways that simplify the pursuit of complex tasks and goals (Todd & Manaligod 2017, 123). The crucial idea, here, is two-fold: the structure of an agent's priority map is shaped by their history, broadly construed; and the topography of this map determines which features of the world are likely to attract or repel attention. This is why combat veterans returning from Afghanistan tend to prioritize combat-related stimuli (Todd et al., 2015b); it is why passengers on a flight that barely avoided crashing in the Atlantic Ocean continue to show "attentional tuning to stimuli associated with the crash years after the event" (Lee et al 2013; cited in Todd & Manaligod 2017, 124); and more mundanely, it is why interpersonal relationships often fall into habitual patterns of misunderstanding, which are grounded in past interactions and their evaluative salience. In each case, attentional parameters are adjusted to yield a stable picture of the world, which highlights the aspects of an environment that are likely to be relevant to current and ongoing projects; this picture of the world then shapes both task-based and feature-based forms of attentional salience (Kryklyvyy & Todd 2018).

All told, the implementation of priority maps is likely to be quite complex, as attentional salience is shaped by numerous interacting processes, operating over different properties of the world, and different time-scales. Some forms of attentional mapping may be implemented by geometric or relational maps in the medial temporal lobe, the parietal cortex, or the superior colliculus. But in some cases, visual activity might be modulated more directly by circuits in the amygdala and the locus coeruleus, on the basis of implicit attentional sets that are attuned to affectively and motivationally salient stimuli (Todd et al., 2012; Todd & Manaligod 2017, 127). Given its sensitivity to contextual information, norepinephrine circuits in the locus corellicus are also likely to "play a role in contextualizing the sources of salience that are prioritized in any given state space" (Todd & Manaligod 2017, 128); this hypothesis gains support from data showing that a common genetic variation that affects the availability of norepinephrine (the deletion variant of *ADRA2b*) profoundly impacts the experience of affectively salient stimuli, leading to more vivid emotional experience (Todd et al 2015a). Some of these processes

operate rapidly and directly on neural structures in the early visual cortex, biasing competition for selection in favor of things that are affectively or motivationally salient; others operate over longer time-scales, shaping the consolidation of information, as well as the subcortical mechanisms that guide our behavior. But collectively they shape a priority map, which orients attention, and structures patterns of thought and behavior.

Finally, Todd and her colleagues argue that genetic predispositions can shape a priority map by modulating the attentional salience of affectively salient stimuli (Todd et al 2013). Just as importantly, they acknowledge that early experience can make certain phenomena seem more salient, as can repeated engagements with specific stimuli—and these facts become highly salient when we look at patterns of attention that emerge outside of the lab, as I argue in Section 4. But even in these cases, an agent's priority map of the world will be continuously shaped by their actions, and by the situations where they find themselves (Todd & Manaligod 2017, 133). For example, individual commitments can restructure the topography of a priority map, allowing a person to orient toward politically salient information, and to avoid information that contradicts their current goals and values (Whitman et al 2018). Explicitly represented goals can shape what we pay attention to for a short time, though they will always compete with the more implicit forms of habitual attention (Jiang 2017). And the topography of a priority map can even shift to align with our current needs and interests—which are shaped by our memories, and realized by distributed patterns of neural reactivation (Todd & Manaligod 2017, 123). This is important, as the things that should be salient to us when we walk around an unfamiliar city are different from the things that should be salient to us when we walk along a familiar path to the office or a café. But at the same time, many of our attentional strategies will drift toward statistically stable and evaluatively significant aspects of the world as we encounter it, even if different attentional sets will be highlighted in different contexts.

The upshot of this neo-Sellarsian account of picturing is that we possess non-conceptual, but highly structured priority maps, which shape our patterns of attention, and which guide goal directed behavior.[7] In line with Sellars's argument, I have suggested that our most basic ways of tracking and responding to the structure of our environment are implemented by priority maps, which are constructed through a process that relies on some endogenous constraints, as well as some forms of reward-driven and statistical learning. These mechanisms operate mechanically, through feedback-driven learning; and they are shaped by a person's practical engagements with the world. Consequently, they yield embodied and situated capacities to act in

7 I contend that these sub-personal processes, which operate through patterns of attunement, are unlikely to carve the world in ways that map the categorical structure of conceptual thought. For further discussions of this complex issue see Huebner (in press), which includes a further discussion of the example that I explore in the next section.

ways that are highly responsive to regularities in the environment, and they do so without relying on conceptually structured thoughts (thereby paralleling Nutmeg's ability to act responsively, even though she lacks complex beliefs about the world). But humans live in linguistically structured environments; and their priority maps can represent semantically intelligible contents. These representations will often become relatively stable aspects of a person's way of representing the world, which reflect their ability to describe things in conceptual terms. And these conceptually structured thoughts will often shape the way that typically functioning people encounter the world. To understand how these kinds of capacities arise, we need to turn to a more robust theory of signifying, which explains how word-and-sentence shaped thoughts are implemented within the framework of non-conceptual picturing that I have developed thus far (cf., Levine 2007, 254).

3. How do concepts signify?

In contrast to the standard assumption that conceptual states acquire the content that they have through causal contact with the world, Sellars argues that what a term signifies is a matter of functional classification, which is situated within a broader network of social and historical practices. He contends that learning what a word signifies is a matter of learning how to use that word in various contexts, in ways that accord with local inferential practices; though he also acknowledges that complex networks of causal relations sustain "linguistic behavior both in its own internal patterns and in its relationship to entities in the world" (O'Shea 2007). Though Sellars never puts the point exactly this way, we can treat meaningful uses of language as part of a typical person's *habitus,* and we can treat signifying as an activity that depends on the use of parts of a cognitive map to actualize capacities for rational thought and behavior. This is an interesting claim, which pushes us toward a novel way of understanding conceptual content. And to see what this amounts to, it will help to first examine how meaningful representations operate across different languages, before returning to claims about the nature mental representation.[8]

8 How is the approach I develop related to more familiar discussions of semantics? To begin with, I hold that there are formal and logical constraints on sentence formation, which are acquired by way of a socially situated learning process. So I accept a form of realism and internalism about formal semantics, where meaning is constrained by hierarchical structures that are interpretable by a system that implements what linguists typically call Logical Form (LF). But we cannot infer much about the worldview of a speaker from the grammar of their language (Hale 1986). So I contend that the resulting logically-structured expressions will always be underspecified, and that hearers will have to infer an understanding of the similarities and differences between their picture of the world and the speaker's. This requires adopting an interpretivist lexical semantics, where meaning is a matter of functional classification. Like Sellars, I thus hold that assumptions about meaning will tend to track similarities in the ways that linguistic terms are used

When I learn from an Arabic speaking friend that 'قوهق (*qahwah*) means coffee', I learn that their use of 'قوهق' plays the same functional role as my use of 'coffee'. I don't learn anything new about the meaning of 'coffee'. But I do gain a sense of how they are likely to use 'قوهق'. Of course, there will be numerous differences in the ways that we write and speak the relevant word; and there will be differences in the ways that we picture the relevant beverage. But in learning that 'قوهق means coffee', we open up space to discuss salient aspects of the delicious beverage, and to exchange reasons for including or excluding more marginal substances (e.g., Is Dunkin' Donuts cold brew really coffee? How about a shot of Starbucks espresso in 250 ml of milk?). Given our unique social and historical situations, we will each possess a picture of the world that has been partly caused by encounters with coffee. And in general, there will be structural relations between the parts of our pictures which characterize the space of possible coffee experiences. But we can't access these aspects of our priority maps directly, any more than we could access the recording on a vinyl record without a turntable. Nonetheless, a long conversation would allow us to explore a wide range of inferences that we were both willing to assent to with regard to 'coffee'; and a shorter discussion might reveal many shared ways of picturing coffee.

The structure of a person's picture of the world will be at least partially reflected in their linguistic behavior, since there is a tight causal connection between picturing and behaving. And in some cases, the mapping between neural activity and linguistic behavior can even underwrite ways of treating patterns of neural activity "as symbols which have meaning, which belong to the order of signification" (Sellars 1960, §44). But doing so is never straightforward. After all, a person's picture of coffee-relevant phenomena will shift and develop as they move around the world, learning more about the use of 'coffee', and more about the relevant substance. Moreover, differences in the coffee-relevant parts of a cognitive map will emerge as the result of differences in people's learning histories, as well as difference in the evaluative salience of coffee. So some people will come to represent coffee (still non-conceptually) in a highly abstract way, as the black stuff that they tend to drink in the morning; while others will develop a more detailed picture, which is structured around knowledge of coffee beans, differences in terroir, and differences in techniques for roasting beans and extracting coffee from them.[9] But just as importantly, we should find that a person's use of 'coffee'

(with usage being determined by something like language-entry rules, intralinguistic transitions, and language-exit rules). To unify these two perspectives, we need something very much like the account of semantic content advanced by Donald Davidson (1986). I currently believe that a promising explanation of how we interpret meaningful claims is likely to approximate Elin McCready's formal model of emotive meaning and dog whistles, which demonstrates how speakers coordinate to extract meaning from underspecified representations (McCready 2012; Henderson & McCready in press). Over the course of this section, I attempt to unpack this complex set of claims.

9 Following Tolman (1948, 193), we might say that some maps of the coffee domain are narrow and strip-like, while others are broad and comprehensive. Both types of maps can

can expand or contract to fit different contexts, just as their priority map can change across different contexts. I am likely to be more discriminating when I am talking to friends who are baristas, or coffee connoisseurs; and I am likely to become less discerning as I spend more time with people who couldn't care less about the caffeine containing beverage that they drink. But in any case, the process of learning, storing, and using the parts of pictures for particular purposes can open up the possibility of treating those mental states that are used in the way that I typically use the word 'coffee'; and this is what underwrites the claim that a particular pattern of neural activity signifies 'coffee' (Sellars 1960, §52).

The critical upshot of this discussion is that the dynamic nature of attentional landscapes, and our strategies for picturing the world, make it unlikely that we will find a stable and systematic mental language, though points of stability are likely to emerge in attentional landscapes as people learn to speak a language.[10] Put much too simply, learning to use parts of non-conceptual pictures for thinking, planning, and remembering, requires learned and habitual tendencies to use 'coffee' to label the relevant parts of a picture (Clark 1996). When we turn to questions about mental representation, we must therefore distinguish: 1) the ability to use part of a picture to orient toward coffee-relevant aspects of the world from 2) the ability to label part of a picture with the relevant linguistic concept.[11] Both processes are implemented mechanically. And both are necessary to sustain the functional mapping between the current state of a system, and the everyday use of a linguistic term. This means that conceptual thought is stabilized through the construction and use of a picture of the world, which interacts with our ability to treat aspects of a picture as a meaningful representation of the world, and to use this labelled representation for thought and communication. But it is worth dwelling on these points, as it's easy to miss their importance.

Recall that picturing is a non-conceptual capacity, which orients us toward particular aspects of the world, and serves as the model that we can employ in planning and deliberating. Within such a model, every coffee-

be correct, as far as they go; and both can be used to guide relevant forms of actions. But broad and comprehensive maps allow for a wider range of inferential relations, about a wider range of different substances. Importantly, people with different coffee-maps will be able to have a discussion of coffee, even where there are robust differences in the ways that they represent coffee.

10 This is not an anti-nativist commitment, nor even an argument against Universal Grammar. While I will not defend this claim here, this approach leaves room for a faculty of language that uses statistical learning to produce stable and resilient linguistic properties (see Lightfoot 2017 for a readable defense of this claim).

11 I initially developed this claim in the context of a paper on racial bias, where I approached this link as a difference between the capacities that we possess socially, and the capacities that we possess individually. I no longer think that this is the right way to develop the argument. But for another way of thinking about this distinction, which draws on a predictive coding framework, and which may be closer to Sellars's own commitments, see Sachs (2018).

representation depends on coffee-relevant dispositions, which are sustained by reliable tendencies to picture the world in particular ways (Levine 2007, 254). And the aspects of a picture that are coffee-relevant can change over time, as new experiences become relevant, and as old ones become irrelevant. In my own case, 'coffee' thoughts are organized around a network of specific interests (e.g., a desire for caffeine, a desire for a particular taste, and my enjoyment of specific flavor profiles), which orient me toward the construction of a cognitive map that will lead me to pursue, brew, drink, and think about coffee across a wide range of different contexts. These interests allow me to ignore features of coffee-drinking experiences that are irrelevant to future coffee-seeking behavior. And they allow me to orient toward those features of coffee drinking that improve my understanding of the diversity and complexity of the coffee domain. When I try a bean from a new roastery, or from a new region, I may focus on the flavor profile of the coffee that I drink. And when I visit a new city, I will seek out the most highly recommended places to drink coffee.

Over time, the coffee-relevant aspects of my picture of the world have become relatively stable, as I have habituated to thinking about coffee in particular ways, which accord with my historical and social situation. To the extent that I am more snobbish about coffee than many of my friends, this is the result of encounters that have shaped the way that I picture the role of coffee in the world; and my idiosyncratic way of being attuned to the world provides me with the foundation for my conceptually structured thoughts about coffee. But for Sellars, the storing of a representation can only ever be part of the story. And a plausible account of conceptual states will also need to explain how stored information is "added to, scanned, and responded to by the other components" of the cognitive system (Sellars 1960, §40). So it is just as important to note that most typical humans have the capacity to use linguistic labels to identify aspects of their cognitive map, and to re-identify the things they have learned about. There is some reason to suppose that a neural network that processes linguistic information will recapitulate the constituent structure of linguistic representations.[12] Where a person internalizes linguistic practices, this will yield parts of the cognitive system that are less contextually variable, and this will allow for re-usable 'words' that retain the same structural relations across sentences where they occur (NB: we are in the domain of formal semantics here, and not the domain of lexical semantics). To the extent that the interfaces between this system and the mapping system allow for (at least) momentary points of stability, we will find internal structures that can serve as the targets of signification.

12 Schonbein (2012) defends the relevant conceptual claim by reference to connectionist systems; and Ding et al (2017) show how neural oscillations might sustain the kinds of hierarchical structures that are commonly posited by theories of generative grammar.

These internal structures are likely to remain somewhat idiosyncratic, but they will be shareable, and it will be possible to explore and revise them, precisely because of the way that they integrate stable structure with dynamic contents. So the mapping between a word like 'coffee' and the internal state that signifies 'coffee' will always be complex. But this shouldn't be surprising. Even in the linguistic domain, the mapping between conceptual states is often more complex than the matching of two similar words. The differences in orthography between an Arabic and English word might seem salient to a first time observer; but these languages use a single morpheme to signify 'coffee', and the English term is descended from 'قوهق'. But a conceptual understanding of 'coffee' doesn't require using a mono-morphemic term, and the term that is used doesn't need to be a descendant of 'قوهق'. For example, an Ojibwe speaker might use 'makade-mashkikiwaaboo' to signify 'coffee'. And while a more literal translation might be something like 'black medicine water', the multi-morphemic structure of this term is unlikely to have a deep effect on their conceptual understanding of the relevant substance. There will be differences in the linguistic structures that they construct; and my use of the mono-morphemic lexicalization 'coffee' my block my ability to use a phrase like 'black medicine water' to express my concept (Poser 1992). But these effects are likely to be generated by the way that syntactic information is stored in the brain; and across many different contexts, we will find that the use of 'makade-mashkikiwaaboo' is similar to may use of 'coffee'. And this is all that must be the case for us to establish a relation of signifying between these two differently structured terms. Pushing further, we might even imagine someone who uses a more complex syntactic construction in the way that I use 'coffee'. While there would be strange lexical implications, someone could even use 'amazing and delicious black energy water of the gods' to signify 'coffee', so long as they reliably used this construction in ways that were conceptually similar to my use of 'coffee'. So long as I can find a way to map my use of 'coffee' onto their use of 'amazing and delicious black energy water of the gods', I can establish similarities between their picture of the world and mine.[13]

This brings us back to the core Sellarsian claim: There are multiple causal routes to the acquisition and use of a term that signifies 'coffee'; and the meaning of 'coffee' can't be specified by appeal to any of them! 'Coffee' means what it does in virtue of its use in practices of functional categorization. For some people, 'coffee' may be little more than a linguistic label, which they have linked to a substance they have read about in books or seen in films; for others the inferential structure of 'coffee' may be connected to many

13 I take this to be a plausible way of redeeming Sellars's claims about dot-quotation in terms of linguistic mappings. Thanks to Carl Sachs for noticing that this is what I was doing, and for suggesting that I leave further development of this argument for a subsequent paper. As the previous footnote suggests, there is a lot of work to do here.

experiences of drinking, brewing, or growing coffee, which structure their sense of what coffee is, and what coffee can do. And while this leads to patterns of difference in the use of 'coffee' (or whatever term signifies 'coffee'), there will often be enough similarity between two people to sustain practices of giving and asking for reasons. From here, ongoing forms of learning and behavior-shaping can draw people closer to one another, both in the way that they picture the world, and in the ways that they use specific terms for thought and communication. There will always be differences in the class of inferences that two people will draw about 'coffee'; but to the extent that these differences are situated among a broader range of similarities, the presence of such states and capacities will allow for practices of communication that can revise and reshape a person's picture of the world.

4. Socialized attention and distorted maps

My primary claim so far is that people develop strategies for navigating their social world by learning to track the things that are most salient to them. This has important implications for the way that they attend to different features of the world; and it has important implications for the ways that they learn to categorize, and to think conceptually. On the one hand, people "learn to navigate the world by attending to the predictability and frequency of objects and events, their meanings in relation to each other, and their associations with reward and punishment" (Todd & Manaligod 2017, 122–123); this yields dynamically structured models that allow people to figure out how to get around in the world. On the other hand, these models can shape the conceptually articulated thoughts that a person will entertain, as well as the kinds of inferences they are willing to carry out. When these capacities are integrated into a single perspective, they yield typically human ways of picturing the world, which may feel—from the inside—as though they are conceptually organized. Perhaps this is why philosophers have often characterized the human mind in ways that appeal to a Language of Thought, which is organized around "word-sized concepts, sentence-sized intentional states and argument-sized inferences" (Williams 2018, 153). But if my argument is roughly correct, our capacity for conceptually-structured thought is an artifact of our socially situated nature, and the concepts we employ reflect the categories that are salient to the people we interact with. My aim in this section and the sequel is to show that this fact has significant implications for the study of situated forms of social cognition.

4.1 Socially sculpted attention

Let's begin with the growing range of evidence that our understanding of the social world arises through active and ongoing participation in culturally scripted patterns of behavior, which lead to the development of "attention allocation strategies that are consistent with local cultural assumptions" (Park

& Kitayama 2011, 77). In line with the approach I developed in Section 2, this appears to yield priority maps that are shaped by rewards that are received for acting in accordance with local norms, and by criticisms that are received for acting in ways that are socially deviant. Recent data also suggests that we can learn how to think about the nature of social groups by tracking how people interact with one another (Lau et al in press); and here too, there is reason to believe that many of the same reward-driven mechanisms are at play (Klucharev et al 2009, 2011). But no matter how these categories are learned, the actions that people take, and the conversations that they have with one another will shape the structure of the world that they and others inhabit—and this will produce stable patterns in the attitudes and attentional strategies that people acquire as socially situated agents (Kitayama & Uskul 2011, 422). This is just to say that both priority maps and conceptual representations are shaped by ongoing social feedback. For example, as they discuss things that are important, they will tend to converge on shared representations of past events (Hirst et al 2018). And this can even lead to the social suppression of aversive information, and to highlighting positive information; in ways that will produce individual memories that are anchored to the groups that people are part of (Coman & Berry 2015; Coman & Hirst 2012, 2015). These effects on memory, and their effects on priority maps deserve further discussion; and I hope to return to them in a future paper. But for now, I want to turn to the ways in which individual differences in learning histories can yield divergent ways of picturing the social world

Numerous studies have revealed that White, middle class, North Americans tend to converge on attentional strategies that insulate their thinking from contextual factors, and focus their attention on discrete entities; they tend to see individual merit as salient, and contextual factors as background phenomena (Adams et al 2010; Markus & Kitayama 2010; Park & Kitayama 2011). This is not an innate disposition, and it's not a fact about every middle class White American. But it's stable pattern of attunement to the kinds of things that such people typically encounter. Attentional maps that highlight individual achievements are reinforced through practices of praising and blaming individuals for their actions; and they are scaffolded by ongoing engagements with "mobility affording transportation and communication infrastructure, the practice of 'leaving home' in young adulthood, the daily practice of eating from individual place settings, and residence in self contained apartment units" (Adams et al 2010, 283). By routinely experiencing these kinds of concrete social realities and social opportunities, people develop a picture of the world where "exploration, expression, and indulgence of unique, individual feelings" are the primary goods to be pursued (Adams et al 2010, 284). And this typically leads them to "express a desire for mastery, control, achievement, choice, self-expression, or uniqueness" (Markus & Kitayama 2010, 421). These habits and dispositions are ways of picturing a categorically structured world, which is organized

by relationships of individual success and failures, and which yields the expectation that self-interested actions will tend to yield such success (Markus & Kitayama 2010, 428).

Converging data from Michael Kraus and his colleagues (2012) suggest that economic and social constraints can also shape an agent's way of picturing the world. They argue that the ongoing experience of economic and social freedom leads to the development of cognitive strategies that are focused on internal states, and that treat these states as the dominant influence on thought and behavior. When people are chronically immersed in environments of relative abundance and elevated social rank, they "are free to pursue the goals and interests they choose for themselves", and they can "pursue these goals and interests relatively free of concerns about their material costs" (Kraus et al 2012, 550). And as a result, middle class and upper class individuals tend to prioritize individualized selves, and assume that behavior is generally caused by individuals, instead of depending on contextual or situational factors. By contrast, where the stability of necessary resources is uncertain and unpredictable, this can lead to the attentional prioritization of contextual and situational factors. So people who live in less stable neighborhoods, who face constant economic instability, and who depend on constantly fluctuating institutional resources often experience the world as socially structured, institutionally constrained, and more limited in social opportunities (Kraus et al 2012, 549). As a result, people who are chronically immersed in these sorts of environments tend to develop attentional strategies that are sensitive to cases of overt social control, and they tend to be aware of the continual recording of their actions in accordance with the preferred frameworks of people in positions of social power.

4.2 Racially sculpted attention

My argument can be summarized as follows. Our habits of attention are shaped by ongoing patterns of feedback, from the people that we interact with, and from our movements through the social world; the ongoing reshaping of our priority maps makes it possible for us to acquire skills that we need to succeed in our local environments; and once we have learned to track all of the relevant social phenomena, our attentional biases can help to minimize the amount of cognitive effort that is required to act in a socially accepted way. This can often be a good thing. But where patterns of exclusion and oppression become entrenched in the material and ideological structures of our cities and social spaces, this same process can yield distorted pictures of the world, which nonetheless *feel like* they represent the world 'as it is'. For example, in contexts where people are presented with racialized imagery in films, novels, and news sources, patterns of attentional salience will begin to stabilize around these aspects of their experience. And this will lead them to orient toward any information that is consistent with this racially structured

priority map; but just as importantly, they will tend to suppress information that is at odds with their priority map.

Imagine someone is walking down an unfamiliar alley in an unfamiliar city. A friend has told them that this is a dangerous place to be. And as they look around, they notice familiar markers of threat and danger. Or at least that is how things seem to them. Perhaps they have seen similar things in contexts where they felt scared; or perhaps they merely encountered tales of similar situations in novels, films, or news sources that were saturated with danger and violence. But in any case, they become more aware of their surroundings; they search for lurking threats; and they ruminate on potential encounters with danger. Their muscles grow tense, their heart begins to race, and their rate of respiration increases—all in the service of preparing to manage the expected danger. If this person had never *learned* that this was a potentially dangerous situation, things would probably be quite different. But they possess a robust picture of the world, which is anchored to their learning history, and to past experiences; and this picture affects their threshold for experiencing fear. This might be a good thing, if it helps them avoid a genuine threat to their wellbeing. It might also send them running from a harmless rat that scurries from behind a trash bin. And it may lead them to act on racist or classist biases, causing substantial harm to innocent individuals.

To make this set of claims more concrete, consider the factors that are at play in first-person shooter tasks (FPST), where participants are asked to decide whether someone is holding a gun or an innocuous object (e.g., a cell phone or a wallet), and to use a button press to respond (shoot vs don't shoot). In carrying out this type of task, people must integrate multiple sources of information; some of the relevant information will be conceptual, some will be affectively structured, and some will be organized by habituated expectations (for similar claims in the context of implicit bias, see Amodio 2014; Faucher & Poirer 2017, Van Bavel et al 2012). And the way that these sources of information are integrated will have an impact on the patterns of response that people tend to display in this kind of task.[14] In a meta-analysis of 42 experiments, Yara Mekawi & Konrad Bresin (2015, 124) found that "participants were quicker to shoot armed Black ['suspects], slower to not shoot unarmed Black ['suspects], and were more likely to have a liberal shooting threshold for Black ['suspects]". But while people were more likely to respond by choosing to shoot a Black person (vs a White person), there

14 Elsewhere, I have defended a computational approach to implicit bias, which explains how these sources of information are likely to be integrated to yield decisions (Huebner 2016, 2018). The view that I defend shares much in common with Edouard Machery's dispositional approach to implicit cognition. Here, I build on important claims that have emerged in the context of recent discussions about the role of attention in cognitive permeation (e.g., Machery 2016 and Gross 2017), and I extend this approach to cover the nearby phenomena known as 'shooter bias', which draws on more robust models (e.g., signal detection theory), and which has received far less philosophical discussion than 'implicit bias'.

were not significant differences in falsealarm rates. This suggests that people are not seeing innocuous objects as guns when they are in the hands of Black people, even though they are more willing to choose to shoot them. To me, this suggests that the problem runs much deeper, and it tells us something significant about the socially sculpted attentional map of race in the US.[15] But in order to see why I believe that this is the case, we will need to look at data showing which kinds of social information are at play in the production of such responses.

To begin with, there is evidence that patterns of response in experiments using FPST are affected by cultural factors (Mekawi & Bresin 2015). For example, people who live in states with more permissive gun laws tend to be more likely to shoot overall; and people who live in cities with lower proportions of White inhabitants tend to display higher levels of anti-Black bias. Contextual factors also seem to matter a great deal (Correll et al 2011): where 'suspects' are situated amid signals of social threat, shooting thresholds are lower across the board, leading to similar patterns of response to white and Black 'suspects'; likewise, when people read stories about white criminals before a FPST, the topography of their attentional landscape shifts, in ways that lead them to respond to both white and Black 'suspects' as equally threatening; finally, a training period where higher numbers of white people are presented with guns, shifts attention toward the assumption that white 'suspects' are more likely to have a gun. Presumably, the second and third effects are short term artifacts of the experimental environment. But the first effect is likely to depend on a robust positive association between proportion of Black people in a neighborhood and white people's fear of crime (Chiricos, Hogan, & Gertz, 1997). The presence of a high number of Black people in a neighborhood is often sufficient to trigger the assumption that a neighborhood is dirty, disordered, and dangerous (Sampson & Raudenbush 2004). And where people feel like a situation or person is dangerous, their threshold for responding to a potential threat will be greatly reduced.

Even more strikingly, Joshua Correll and his colleagues (2015) have shown that attention is likely to play a critical role in FPSTs. Using an eye tracker, they found a significant effect of race on the visual angle between the object to be categorized (the weapon or the innocuous object) and the fovea at the time of response. More specifically, the visual angle was larger at the point where a decision was made about whether to shoot a Black person (Black,

15 In a study that monitored event-related potentials (ERPs) during a FPST task, Correll et al (2006) found that a larger P200 response to images of Black people predicted the extent to which people were quicker to choose to "shoot" armed Black 'suspects', and to "not shoot" unarmed white 'suspects'. Similarly, Amodio et al (2004) found larger event-related negativity (which they interpret as activity in the anterior cinculate cortex) where race and object are stereotypically incongruent, suggesting perceived conflict; moreover, they found that more pronounced event related negativity (ERN) correlated with greater accuracy and slower reaction times, suggesting that increased cognitive control could be used to inhibit this prepotent response.

M=2.08°; White, M=1.59°); and this difference persisted in contexts where a weapon was present (Black, M=2.03°; White, M=1.41°). This suggests that people need more precise visual information to decide whether to shoot a white person, and more precise information to see whether they are holding a gun. This is what we should expect if most people possess a priority map that highlights Black people as threatening. Put much too simply, people are more likely to choose to shoot a Black person because their attentional landscape highlights the connection between Black people and danger; so they need less visual information to decide in favor of the hypothesis that a Black person is dangerous because they possess a picture of the world that highlights the connection between race and threat (cf., Machery 2016, 64).

Intriguingly, things look different when police officers take part in FPSTs. They tend to be quicker, more accurate, and more sensitive to the presence of guns (Correll 2007). This makes sense, as they are typically trained to hold their fire when they are uncertain; and they often cultivate forms of reflexive control, using training conditions where people fire paintballs or simulated and painful ammo. And as a result, they should be more likely to ignore irrelevant information, and to focus on situationally relevant stimuli. Even so, police officers tend to be slower to respond to counter-stereotype situations; and those who work in communities where there are larger Black populations, and higher levels of violent crime, tend to have more biased response latencies. And crucially where their work environment reinforces the salience of the connection between race and violence, high levels of training are insufficient to mitigate the effects of racial bias. Like untrained community members, officers who work in Gang Units and Violent Crime units are more likely to choose to shoot Black than White 'suspects' in a first person shooter task (Sim et al 2013, 300).

Of course, these kinds of tasks are not ecologically valid, and it is difficult to know what to infer from FPST tasks. Indeed, a recent experiment using a more realistic and more immersive simulation, where people had to decide whether to shoot a laser-equipped handgun, found that people were more likely to 'shoot' unarmed white 'suspects' than unarmed Black 'suspects (46/184 vs 1/47; James et al 2014). These data seem to contradict the data that have been collected on computers using FPST tasks. But in thinking about this experiment, is important to note that this was a task that required a decision to shoot or not, whereas standard FPSTs require a decision about whether to push a button on the left or the right. And critically, participants in this task took longer to decide whether or not to shoot Black 'suspects' (even when they were armed); and data collected using electroencephalogram (EEG) during the simulation revealed higher levels of alpha-wave suppression for armed as well as unarmed Black 'suspects'. There are multiple ways of interpreting these data, but the most plausible hypothesis is that Black 'suspects' were always perceived as threatening, that the alpha-wave suppression reveals an inhibitory response, and that the slower response

reveals active suppression of fear as a result of the desire not to appear racist. This strikes me as plausible, because the goal of not appearing racist is likely to play a much more prominent role in shaping a decision about whether to act or refrain from acting.

It would require a more substantial argument to establish this claim conclusively. But so far as I can tell, it accords with all of the existing data, as well as the argument I have developed throughout this paper. More importantly, it suggests that if we could always trust people to inhibit their initial responses, it would be possible to reshape problematic patterns of socially entrenched behavior merely by cultivating these kinds of goals. Unfortunately, I'm not optimistic that these data provide insights into human behavior outside of laboratory environments. To prevent the emergence of racial bias, people would need to cultivate highly salient and conceptually structured goals of inhibiting racially charged responses, and this goal would need to play an ongoing role in real life decisions. I doubt that this is likely to be the dominant motivation in real-life situations for most people; and in the absence of strong anti-racist motivations, and active attempts to suppress unexpected and problematic responses, we have little reason to believe that these kinds of attentional effects will generalize to the kinds of situations that we should be concerned about (cf., Huebner 2016).

4.3 The production of racially sculpted concepts

With this background in hand, I want to turn in closing to a suggestion about how racialized concepts become stable in contexts where they are experienced as highly salient. To begin with, note that people in the United States tend to overestimate the percentage of the population who are Black, Jewish, Asian, and Latinx; people from communities with larger white populations tend to guess that the nation has a larger white population; and people from communities with larger Black populations tend to guess that the nation has a larger Black population (Wong 2007, 401). But shifts in the evaluative salience of different groups can complicate this situation in troubling ways. As Charles Gallagher (2003) argues, people who live in communities where white people interact primarily with other white people (which is the norm in the US) will often acquire highly distorted pictures of the national population. He focuses on three situations that yield the salient distortions.

1. Some people who live in predominantly white spaces encounter racialized minorities primarily in the context of films and news media that present the Black population as dangerous; this heightens their experience of racial anxiety, increasing the salience of both real and virtual encounters with Black people, and causing something like oversampling in perception and memory. Strikingly, Gallagher's data show that this can lead to extreme over-estimations of the Black population (40–60%; actual 12%).

2. Other people who encounter racialized minorities find collective demands for racial justice to be more salient. This can focus attention on presentations of Black activists in the news media, and on social situations where demands for change are being made; and here too, people tend to substantially overestimate the percentage of the US population that is Black. And just as importantly, it can lead them to make illicit assumptions about Black people being better off than white people in achieving their needs and interests.
3. Finally, in contexts where people become anxious about demographic shifts that could make the US a majority-minority nation, this can heighten the salience of encounters with Black people, and again lead to a situation where people will tend to overestimate the proportion of the population that is Black.

These specific effects are distinctive of the US, where a specific history of racial injustice makes race a highly salient feature of the social environment. However, similar effects are likely to emerge in any context where there are interactions between statistical stabilities and socially reinforced evaluations, and wherever the salience of a group recruits attention, shapes memories, and affects the decisions that people make. Given the coalitional nature of human psychology, this means that similar kinds of phenomena are likely to show up across most populations (Van Bavel & Pereira 2018). One place where this affect is obvious is on the kinds of assumptions and inferences that are beginning to emerge and solidify around issues of immigration, in many parts of the world. Paralleling my discussion about 'coffee' above, there are multiple routes to the acquisition and use of a socially salient term like 'immigrant' and the meaning of these terms can't be specified by appeal to any of these causal relations. Terms like 'immigrant', along with countless other terms, mean what they do in virtue of their use in practices of functional categorization; but social forces can lead to the emergence of highly divergent ways of using such a term, which often become clear only after sustained discussion, which often becomes highly unpleasant. For some people, 'immigrant' may be little more than a linguistic label, which is anchored to things they've heard on talk radio, seen in the news, or read on a presidential twitter feed; or it might derive from the inflammatory rhetoric of a political campaign. Such people will become more attentive to information that accords with this acquired picture of the world, and their thoughts about immigration will be driven by anxiety. For other people, the inferential structure of 'immigrant' may be more highly elaborated, as it may be shaped by experiences with friends, colleagues, or random people throughout the city; they will have a richer understanding of why people move to a new country, as well as a more robust understanding of the roles that immigrants play in a vibrant and thriving society. These kinds of differences can produce highly divergent ways of using a term like 'immigrant'. The fact that we use the same word, however,

can often lead us to believe that we picture the world in the same way. And this can lead us to neglect the causal and structural forces that sustain habits of thought and attention.

5. Concluding thoughts

If the argument that I have developed is roughly right, agents learn to situate themselves within the world *they encounter* (cf., Sellars 1978 §28–29). And their ways of picturing the world shape the storage and retrieval of memories, leading to patterns of thought and behavior that are socially shaped and sustained. But local patterns of attunement can yield global patterns of distortion. And where they do:

> We compare, struggle, and wonder how to let go of our personal, subjective view and arrive at an objective recognition of things. We want to be directly in touch with the reality of the world. Yet the objective reality we think exists independently of our sense perceptions is itself a creation of collective consciousness. Our ideas of happiness and suffering, beauty and ugliness, are reflections of the ideas of many people (Thích Nhất Hạnh 2006, 39).

This is not something that is unique to specific ways of encountering the world: all typically functioning humans will develop strategies for prioritizing information, in accordance with their long and short-term goals (Todd & Manaligod 2017, 122). And given the socially structured nature of our goals, they will all tend to operate from within their own pictures of the world. Even so, the fact that these ways of thinking are constructed means that it's possible to reshape what is salient, and to reorient our habituated tendencies to act in specific ways (see Cikara and Van Bavel 2014 for a review). Even the most socially-entrenched biases become less pronounced in contexts where an alternative way of categorizing is available. For example, where people attend to the shared love of a football team, or to shared commitments to a university, their attentional biases shift toward these categories (Van Bavel & Cunningham, 2009), at least so long as these categories are the contextually salient way of dividing up the world. When we look at the world from a different perspective, the features that are most salient will begin to shift. The key question is: what would it take to make such a shift permanent?

Unfortunately, I'm not sure that this is possible. Though there is plenty of room for empirical research on ways of re-shaping these kinds of habituated attentional biases. I believe that the most promising option will require acknowledging the deep respects in which all human interests are interconnected and interdependent. This point is defended in some parts of Buddhist philosophy; and it is nicely summed up in an Ashanti metaphor about a crocodile with two heads, which are fighting with one another

over access to food (Wiredu 1995, 57). If the two heads were to recognize that any food that either of them ate would end up in the same stomach, then the motivation to compete would dissipate, and it would be replaced by cooperative drives for mutual aid and mutual support. According to Ashanti tradition, human conflict can always be reconciled through patterns of dialogue, which highlight shared needs and shared interests. So long as everyone listens, and so long as they all work to build a shared understanding of the situation, the drive toward mutual aid and mutual support will always arise. Of course, dialogue across deep cultural divides is never easy. And in many cases, we get caught up in attempting to defend our own beliefs, without listening to the things that other people need. This occurs both in the context of interpersonal relations, and in the context of cultural differences. When we become anxious, it is harder to be vulnerable, and it is harder to see points where new paths forward can be developed. There is evidence that feelings of racial anxiety can trigger an increase in the release of norepinephrine, which can compromise cognitive control, and lead to forms of thought and behavior that are more directly shaped by the implicit structure of priority maps (cf., Amodio et al 2004; Godsil & Richardson 2016). As I noted above, differences in the availability of norepinephrine can shift the affective salience of different kinds of stimuli, and they can shape what kinds of features of the world we attend to. So we need to find ways of mitigating anxiety; and this will require either contemplative training, or prefigurative forms of social practice (and maybe both). But that's another story for another day.

6. Works cited:

Abraham, T. (2019). Cybernetics. In *The Routledge Handbook of the Computational Mind*. M. Sprevak & M. Colombo, eds. Routledge.

Adams, G., Salter, P. S., Pickett, K. M., Kurtis, T., & Phillips, N. L. (2010). Behavior as mind in context. *The mind in context*, 277–306.

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences* 1191, 42–61.

Akins, K. (1996). Of sensory systems and the "aboutness" of mental states. *The Journal of Philosophy*, *93*(7), 337–372.

Amodio, D. M. (2014). Dual Experiences, Multiple Processes: Looking Beyond Dualities for Mechanisms of the Mind. In J. S. Sherman, B. Gawronski & Y. Trope (eds.), *Dual Process Theories of the Social Mind*, NY: Guilford Press, 560–576.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, 15(2), 88–93.

Anderson, B. (2016). The attention habit: How reward learning shapes attentional selection. *Annals of New York Academy Sciences*, 1369 (1), 24–39.

Awh, E., Belopolsky, A.V. &Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16 (8), 437–443.

Chiricos, T., Hogan, M., & Gertz, M. (1997). Racial composition of neighborhood and fear of crime. *Criminology*, *35*(1), 107–132.

Chomsky, N. (1995). Language and nature. *Mind*, *104*(413), 1–61.

Cikara, M., & Van Bavel, J. J. (2014). The neuroscience of intergroup relations: an integrative review. *Perspectives on Psychological Science*, 9(3), 245–274.

Clark, A. (1996). Linguistic anchors in the sea of thought?. *Pragmatics & Cognition*, 4(1), 93–103.

Clark, A. (2014). *Mindware*. Second Edition. Oxford: Oxford University Press.

Coman, A., & Berry, J. N. (2015). Infectious Cognition: Risk perception affects socially shared retrieval-induced forgetting of medical information. *Psychological Science*, 26(12), 1965–1971.

Coman, A., & Hirst, W. (2012). Cognition through a social network. The propagation of induced forgetting and practice effects. *Journal of Experimental Psychology: General*, 141(2), 321–33.

Coman, A., & Hirst, W. (2015). Social identity and socially shared retrieval-induced forgetting: The effects of group membership. *Journal of Experimental Psychology: General*, 144(4), 717–722.

Correll, J., Urland, G. R., & Ito, T. A. (2006). Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, *42*(1), 120–128.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: police officers and racial bias in the decision to shoot. *Journal of personality and social psychology*, *92*(6), 1006.

Correll, J., Wittenbrink, B., Park, B., Judd, C. M., & Goyle, A. (2011). Dangerous enough: Moderating racial bias with contextual threat cues. *Journal of experimental social psychology*, *47*(1), 184–189.

Correll, J., Wittenbrink, B., Crawford, M. T., & Sadler, M. S. (2015). Stereotypic vision: How stereotypes disambiguate visual stimuli. *Journal of personality and social psychology*, 108(2), 219.

Craik, K. (1967). *The nature of explanation. 1943*. Cambridge University, Cambridge UK.

Crockett, M. (2013). Models of morality. *Trends in Cognitive Science*, 17, 8, 363–6.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, 25(5), 736–760.

Cushman, F. (2013). Action, outcome and value: A dual-system framework for morality. *Personality and Social Psychology Review*, 17 (3), 273–292.

Davidson, D. (1986). A nice derangement of epitaphs. *Philosophical grounds of rationality: Intentions, categories, ends*, *4*, 157–174.

De Brigard, F., Brady, T. F., Ruzic, L., & Schacter, D. L. (2017). Tracking the emergence of memories: A category-learning paradigm to explore schema-driven recognition. *Memory & cognition*, 45(1), 105–120.

De Brigard, F., Hanna, E., St Jacques, P. L., & Schacter, D. L. (2018). How thinking about what could have been affects how we feel about what was. *Cognition and Emotion*, 1–14.

Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron*, *59*(5), 829–838.

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in human neuroscience*, *11*, 481.

Egan, F. (2014). How to think about mental content. *Philosophical Studies*, *170*(1), 115–135.

Egan, F. (2019). The nature and function of content in computational models. In *The Routledge Handbook of the Computational Mind*. M. Sprevak & M. Colombo, eds. Routledge.

Faucher, L. & Poirer, P. (2017). Mother culture, meet mother nature. In Huebner, B. (Ed.). (2017). *The Philosophy of Daniel Dennett*. Oxford University Press.

Gallagher, C. A. (2003). Miscounting race: Explaining Whites' misperceptions of racial group size. *Sociological Perspectives*, *46*(3), 381–396.

Gardner, M. P. H., Schoenbaum, G., & Gershman, S. J. (submitted). Rethinking dopamine prediction errors.

Gershman, S. J. (in press). Uncertainty and exploration. *Decision*.

Godsil, R. D., & Richardson, L. S. (2016). Racial Anxiety. *Iowa L. Rev.*, *102*, 2235.

Gross, S. (2017). Cognitive penetration and attention. *Frontiers in psychology*, *8*, 221.

Haas, J. (2018). An empirical solution to the puzzle of weakness of will. *Synthese*, 1–21.

Hale, K. (1986) Notes on World View and Semantic Categories: Some Warlpiri Examples, in *Features and Projections*. P. Muyskens & H. van Riemsdijk (eds). Dordrecht: Foris, 233–54.

Hebb, D. O. (1949). *The organization of behavior: A neurophysiological approach*. Wiley.

Henderson, R. & McCready, E. (in press). How dogwhistles work. *The proceedings of LENLS*.

Hirst, W., Yamashiro, J. K., & Coman, A. (2018). Collective Memory from a Psychological Perspective. *Trends in Cognitive Sciences*, 22(5), 438–451.

Huebner, B. (2016). Implicit Bias, Reinforcement Learning, and Scaffolded Moral Cognition. In Brownstein, M. & J. Saul (Eds.). *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*. Oxford University Press, 47–79.

Huebner, B. (2018). Reply to Del Pinal and Spaulding. In a symposium on "Conceptual Centrality and Implicit Bias" at *The Brains Blog*. https://goo.gl/2eV3je

Huebner, B. (in press). The interdependence and emptiness of whiteness. In *Buddhism and whiteness*. E. McRae & G. Yancy, eds. Lexington Books.

James, L., Klinger, D., & Vila, B. (2014). Racial and ethnic bias in decisions to shoot seen through a stronger lens: Experimental results from high-fidelity laboratory simulations. *Journal of Experimental Criminology*, 10(3), 323–340.

Jiang, Y. V. (2018). Habitual versus goal-driven attention. *Cortex*, *102*, 107–120.

Kishida, K. T., Saez, I., Lohrenz, T., Witcher, M. R., Laxton, A. W., Tatter, S. B., White, J. P., Ellis, T. L., Phillips, P. E. and Montague, P. R. (2016). Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. *Proceedings of the National Academy of Sciences*, 113(1), 200–205.

Kitayama, S., & Uskul, A. K. (2011). Culture, mind, and the brain: Current evidence and future directions. *Annual review of psychology*, 62, 419–449.

Klucharev V., Hytönen K., Rijpkema M., Smidts A., Fernández G. (2009). Reinforcement learning signal predicts social conformity. *Neuron* 61, 140–151

Klucharev V., Munneke M., Smidts A., Fernández G. (2011). Downregulation of the posterior medial frontal cortex prevents social conformity. *J. Neurosci*. 31, 11934–1194

Kraus, M. W., Piff, P. K., Mendoza-Denton, R., Rheinschmidt, M. L., & Keltner, D. (2012). Social class, solipsism, and contextualism: how the rich are different from the poor. *Psychological review*, 119(3), 54.

Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in cognitive sciences*, 18(9), 457–464.

Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences*, *111*(25), 9079–9084.

Kryklywy, J. H., & Todd, R. M. (2018). Experiential History as a Tuning Parameter for Attention. *Journal of Cognition*, 1(1), 24.

Kukla, R. (2017). Embodied Stances: Realism without Literalism. In *The Philosophy of Daniel Dennett*. B. Huebner, ed. New York: Oxford University Press, 3–31.

Lau, T., Pouncy, H. T., Gershman, S. J., & Cikara, M. (in press). Discovering social groups via latent structure learning. *Journal of Experimental Psychology: General*.

Lee, D., Todd, R. M., Gardhouse, K., Levine, B., & Anderson, A. K. (2013). Enhanced attentional capture in survivors of a single traumatic event. In *Society for Neuroscience Annual Meeting, San Diego, CA, USA*.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, *47*(11), 1940–1951.

Levine, S. M. (2007, September). The place of picturing in Sellars' synoptic vision. In *The Philosophical Forum* (Vol. 38, No. 3, pp. 247–269). Malden, USA: Blackwell Publishing Inc.

Lightfoot, D. (2017). Invariant and variable properties. Inference, 3, 2. Retrieved from http://inference-review.com/article/invariant-and-variable-properties on 31 August 2018.

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review*, 5(4), 356–368.

Machery, E. (2016). Cognitive penetrability: a no-progress report. Zeimbekis, J., & Raftopoulos, A. (Eds), *The cognitive penetrability of perception. New philosophical perspectives*. New York: OUP.

Markus, H. R., & Kitayama, S. (2010). Cultures and selves: A cycle of mutual constitution. *Perspectives on Psychological Science*, 5(4), 420–430.

Maturana, H. R., Lettvin, J. Y., McCulloch, W. S., & Pitts, W. H. (1959). Evidence that cut optic nerve fibers in a frog regenerate to their proper places in the tectum. *Science*, *130*(3390), 1709–1710.

McCready, E. (2012) Emotive equilibria. *Linguistics and Philosophy* 35, 243–283.

Mekawi, Y., & Bresin, K. (2015). Is the evidence from racial bias shooting task studies a smoking gun? Results from a meta-analysis. *Journal of Experimental Social Psychology*, 61, 120–130.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan P. (2012). Computational psychiatry. *Cognitive Science* 16: 72–80.

Nhất Hạnh, T. (2006). *Understanding our mind*. Parallax press.

Newell, A., & Simon, H. (1956). The logic theory machine--A complex information processing system. *IRE Transactions on information theory*, *2*(3), 61–79.

Newell, A. & Simon, H. (1976). Computer Science as Empirical Inquiry: Symbols and Search, *Communications of the ACM*, 19 (3), 113–126

O'Shea, J. (2007). *Wilfrid Sellars: Naturalism with a normative turn*. Polity, Cambridge.

Olen, P. (2018). The Varieties and Origins of Wilfrid Sellars' Behaviorism. In *Sellars and the History of Modern Philosophy*. L. Corti & A. Nunziante, eds. Routledge.

Park, H., & Kitayama, S. (2011). Perceiving through culture: The socialized attention hypothesis. In N. Ambady, K. Nakayama, S. Shimojo and R. B. Adams, Jr. (Eds.), *Social Vision*. New York: Oxford University Press.

Piccinini, G. (2015). *Physical computation: A mechanistic account*. OUP Oxford.

Piccinini, G. & Bahar, S. (2013). Neural Computation and the Computational Theory of Cognition, *Cognitive Science*, 37 (3), 453–488.

Piccinini, G. & Scarantino, A. (2011). Information Processing, Computation, and Cognition, *Journal of Biological Physics*, 37 (1), 1–38.

Thomson, E., & Piccinini, G. (2018). Neural Representations Observed. *Minds and Machines*, *28*(1), 191–235.

Poser, W. J. (1992). Blocking of phrasal constructions by lexical items. *Lexical matters*, 111–130.

Railton, P. (2014). Reliance, Trust, and Belief. *Inquiry*, 57(1), 122–150.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Sachs, C. B. (2018). In Defense of Picturing: Sellars's Philosophy of Mind and Cognitive Neuroscience. https://doi.org/10.1007/s11097–018–9598–3

Salisbury, J. & Schneider, S. (2019). Concepts, symbols and computation: An integrative approach. In *The Routledge Handbook of the Computational Mind*. M. Sprevak & M. Colombo, eds. Routledge.

Sampson, R. J., & Raudenbush, S. W. (2004). Seeing disorder: Neighborhood stigma and the social construction of "broken windows". *Social psychology quarterly*, *67*(4), 319–342.

Schneider, S. (2009). The Nature of Symbols in the Language of Thought, *Mind and Language*, 24, 4, 523–553.

Schonbein, W. (2012). The Linguistic Subversion of Mental Representation. *Minds and Machines*, *22*(3), 235–262.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, 80 (1), 1–27.

Schultz, W. (2010). Dopamine signals for reward value and risk: basic and recent data. *Behavioral and brain functions*, 6(1), 1.

Selfridge, O. (1958). Pandemonium: a paradigm for learning. In *Symposium on the Mechanization of thought Processes,* London: HM Stationery Office.

Sellars, W. (1956). Empiricism and the Philosophy of Mind, in *Minnesota Studies in The Philosophy of Science*, Vol. I, H. Feigl & M. Scriven, eds. Minneapolis: University of Minnesota Press, 253–329.

Sellars, W. (1960). Being and Being Known, *Proceedings of the American Catholic Philosophical Association*, 28–49.

Sellars, W. (1974). Meaning as Functional Classification. *Synthese*, 27, 417–37

Sellars, W. (1978). The Role of Imagination in Kant's Theory of Experience. In *Categories*. H. Johnstone, Jr. (ed.). Pennsylvania State University, 231–45

Sellars, W. (1981). Mental Events. *Philosophical Studies* 39, 325–45.

Shah, A. K., Zhao, J., Mullainathan, S., & Shafir, E. (2018). Money in the mental lives of the poor. *Social Cognition*, 36(1), 4–19.

Sim, J. J., Correll, J., & Sadler, M. S. (2013). Understanding police and expert performance: When training attenuates (vs. exacerbates) stereotypic bias in the decision to shoot. *Personality and social psychology bulletin*, 39(3), 291–304.

Spencer, M. K. (2015). The Category of Habitus: Accidents, Artifacts, and Human Nature. *The Thomist: A Speculative Quarterly Review*, 79(1), 113–154.

Talmi, D., Ziegler, M., Hawksworth, J., Lalani, S., Herman, C. P., & Moscovitch, M. (2013). Emotional stimuli exert parallel effects on attention and memory. *Cognition & emotion*, 27(3), 530–538.

Theeuwes, J. (2018). Visual Selection: Usually Fast and Automatic; Seldom Slow and Volitional. *Journal of Cognition*, 1(1), 21–29.

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Todd, R. M., & Manaligod, M. G. (2017). Implicit guidance of attention: The priority state space framework. *Cortex*, 30(1), e1–8.

Todd, R. M., Müller, D. J., Palombo, D. J., Robertson, A., Eaton, T., Freeman, N., Levine, B., Anderson, A. K. (2013). Deletion variant in the ADRA2B gene increases coupling between emotional responses at encoding and later retrieval of emotional memories. *Neurobiology of Learning and Memory*, 112, 222–229.

Todd, R. M., Cunningham, W. A, Anderson, A. K., & Thompson, E. (2012). Affect-biased attention as emotion regulation. *Trends in Cognitive Sciences*, 16(7), 365–72.

Todd, R. M., Ehlers, M. R., Mueller, D. J., Robertson, A., Freeman, N., Palombo, D. J., Levine, B., & Anderson, A. K. (2015a). Neurogenetic variations in norepinephrine availability enhance perceptual vividness. *Journal of Neuroscience 35* (16), 6506–6516.

Todd, R. M., MacDonald, M. J., Sedge, P., Robertson, A., Jetly, R., Taylor, M. J., & Pang, E. W. (2015). Soldiers with posttraumatic stress disorder see a world full of threat: magnetoencephalography reveals enhanced tuning to combat-related cues. *Biological psychiatry*, *78*(12), 821–829.

Tolman, E. C. (1932). *Purposive behavior in animals and men*. University of California Press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433.

Van Bavel, J. J., & Cunningham, W.A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin*, 35(3), 321–335.

Van Bavel, J. J., J. Xiao & W. A. Cunningham. (2012). Evaluation is a Dynamic Process: Moving Beyond Dual System Models. *Social and Personality Psychology Compass*, 6/6, 438–454.

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An Identity-based model of political belief. *Trends in cognitive sciences*.

Whitman, J. C., Zhao, J., Roberts, K. H., & Todd, R. M. (2018). Political orientation and climate concern shape visual attention to climate change. Climatic Change, 147(3–4), 383–394.

Wiener, N. (1948). *Cybernetics: Control and communication in the animal and the machine*. Wiley.

Williams, D. (2018). Predictive processing and the representation wars. *Minds and Machines*, *28*(1), 141–172.

Wong, C. J. (2007). "Little" and "big" pictures in our heads: Race, local context, and innumeracy about racial groups in the United States. *Public Opinion Quarterly*, *71*(3), 392–412.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, 24(5), 667–677.