Research Statement

Mason Westfall

In my view, the most promising place to understand the mind is the nexus of contemporary cognitive science and our folk psychological concepts. Cognitive science provides our best current understanding of the operations that constitute minds and intelligence, and our folk psychological concepts reveal which psychological distinctions are most relevant to our lives, especially in our interactions with one another. So understood, folk psychology and cognitive science offer complimentary resources to theorizing about the mind. Cognitive science produces our best current understanding of the truthmakers of propositions about the mental, and allows us to more precisely understand how our folk psychological concepts work. Folk psychology offers a more direct interface between the mind and the rest of philosophical theorizing. Often, I think it turns out that the mental distinctions that are relevant to philosophical theorizing and folk psychology are at a higher level of abstraction than models that emerge from cognitive science. This orientation to theorizing about the mind can be applied to many philosophical and cognitive scientific problems. In my work, I have considered problems in the philosophy of perception, other minds, epistemology and the metaphysics of mind. I have also become increasingly interested in what we can learn about natural agency from recent developments in artificial intelligence, and the social epistemology of group interactions.

Philosophy of Perception

In 'Other minds are neither seen nor inferred' (*Synthese*, 2021), I argued that perception-based knowledge of other minds should be understood in terms of what I call 'ampliative perceptual judgments'. These judgments are immediately justified by experience, and yet outstrip what is presented perceptually. So, on my view, other minds are neither seen nor inferred. I reject the common assumption that perception-based knowledge is always either of what is literally perceived, or inferred on the basis of what we perceive. Neither standard option offers a satisfying explanation of perception-based knowledge of mental states. If we simply perceive others' mental states, it's hard to explain why people have taken mental state attribution to pose a distinctive epistemic challenge not found in other cases

of 'high level perception' like recognizing a dog. Many judgments about mental states cannot be inferential though, since we often lack beliefs about the connection between facial expressions and emotions necessary to license the putative inferences.

Ampliative perceptual judgments are plausibly not restricted to mental state attribution. Debates about perceptual learning, the richness of perceptual content, affordances, and aesthetic perception are all often cast in terms that assume the seen–inferred dichotomy is exhaustive. I expect these debates to look interestingly different with ampliative perceptual judgments in mind.

Discussions about perception and social cognition generally focus on whether or not we perceive *mental states*. In 'Perceiving Agency' (*Mind & Language*, 2022), I shift this focus. I mount an empirical argument that we perceive *agency*. Agency perception plays a crucial role in social cognition, functioning to activate higher cognitive forms of social intelligence. When we perceive agents, we begin considering their possible goals, states of knowledge, and so on. As such, agency perception enables us to deploy our social cognition efficiently and, to a first approximation, marks out the limits of the social world.

Ampliative perceptual judgments constrain the epistemology of perception more generally. Many philosophers—prominently Jim Pryor and Declan Smithies—hold that perception justifies beliefs by presenting contents in a special way—'as true' for example. Ampliative perceptual judgments cannot be reconciled with such views. If some judgments are immediately justified by experience, despite experience not presenting content corresponding to them, then perceptual justification cannot be explained by content presentation. Rather, I argue that we should explain perceptual justification in terms of recognitional capacities—an essential and often overlooked component of both ampliative and non-ampliative perceptual judgments. Perceptual justification is grounded in the skillful manifestation of recognitional capacities. A paper making this argument is currently under review.

The Personal, Subpersonal, and Social Cognition

Commonly, in the philosophy of mind and cognitive science, people appeal to a personal–subpersonal distinction. For example, the computations performed by the visual system are generally taken to be 'subpersonal', and hence not properly attributable to the person. By contrast, inferences are taken to be 'personal'—properly attributable to the person. Although this distinction is commonly appealed to, it's puzzling. Considered purely as information processing, many of the computations performed by the visual system are similar to inferences. What's the difference? What makes some aspects of our psychology 'us', but not others?

In 'Constructing persons: On the personal–subpersonal distinction' (*Philosophical Psychology*, 2022), I argue that the personal–subpersonal distinction ought to be understood as an expression of social cognition. Only this view—which I call 'psychological constructionism'—can resolve what I call 'the plurality problem'. The personal level includes states and processes of many different kinds: cognitive states, experiences, inferences, psychological traits, and so on. This plurality is a problem for attempts to explain what makes something personal rather than subpersonal. Recently, though, scholars like Kristin Andrews, Evan Westra and Shannon Spaulding have argued that social cognition is 'pluralistic', meaning that it attributes a plurality of different kinds of states and processes—not just beliefs and desires—in order to achieve our goals, not least predicting and explaining others' behavior. Assuming that they are right, the constructionist can explain and predict the personal level plurality that was so troubling for rival theories.

Supposing constructionism is correct, further questions come into view. Why would social cognition construct a domain of the personal, and what is the structure of social cognition, such that it could serve this function? What, if anything, unifies our pluralistic social cognition? I am endeavoring to answer these questions. I argue that empirical work—especially in developmental psychology—supports the view that social cognition is unified by a proprietary set of computational principles. These principles are defined over the ontological categories of social cognition, e.g. how belief and desire contribute to action, and constitute an explanatory framework that is disjoint from the folk physical framework. As such, we can think of a physical event as *either* situated in the mechanical explanatory framework of folk physics, *or* within the broadly mentalistic framework of social cognition, *but not both* because they are constituted by cognition characterized by independent sets of computational principles. My proposal explains how social cognition can be a cognitive kind, while cutting across more familiar joints in the mind like representational format and modularity, and may even offer some insight into the psychological underpinnings of the mind–body problem. A paper developing this theory is in preparation.

I have also begun to consider constructionism's relevance to other topics in the philosophy of mind. My co-author—Preston Lennon—and I are considering the metaphysics of belief. Why, for example, is it odd to think that someone has a determinate number of beliefs, say 6,845,932? We think it is not just that it would be difficult to count them, like the hairs on your head. Rather, it's *metaphysically* odd. Our explanation is that the truthmakers for belief attributions are varied, including representations in different formats. This slack between belief attributions and their truthmakers admits of indeterminacy. For example, when a subject has both a language-like and a map-like representation of the same information,

it's indeterminate whether that is one belief or two beliefs. Because social cognition is agnostic about format, our constructionist proposal can make good sense of this otherwise puzzling phenomenon, while charting a middle course between Eric Schwitzgebel's dispositionalism, and Eric Mandelbaum and colleagues' psychofunctionalism.

In an invited handbook chapter, I will consider the connections between mindshaping and constructionism. Mindshaping theorists take the fundamental and central function of human social intelligence to be actively influencing one another's minds. Constructionism and a mindshaping approach hold the potential to be mutually illuminating. Insofar as some psychological kinds are constructed, our mindshaping practices contribute to making psychological kinds the kinds that they are, and insofar as the mindshaping thesis is true, a distinctive explanation of why some kinds are constructed is available. Perhaps those kinds are constructed exactly because they are manipulated by our mindshaping practices.

Epistemology and Polarization

I have begun a new research project on the epistemology of polarization. Many commentators, academic and popular, have worried about increasing polarization, especially with respect to political issues. A seductive line of thought takes this situation to be epistemically problematic. I argue, perhaps surprisingly, that more careful attention to detail reveals that polarization as such is not epistemically objectionable. Individuals who become polarized may be fully epistemically rational. The central issue with worrying about polarization epistemically is that polarization is a structural phenomenon, and as such, abstracts away from the content about which subjects' views are changing. But in abstracting away from the content, we lose important distinctions between epistemically virtuous and epistemically vicious polarization. For example, Thi Nguyen's notion of an 'echo chamber', when understood in a non-normative, structural way, equally applies to Q conspiracy theorists and academic climate scientists. Both social-epistemic communities take people espousing putative counter-evidence to the beliefs that characterize their community to thereby be epistemically downgraded, rather than taking the apparent counter-evidence at face value (the essential feature of echo chambers, on Nguyen's view). So I think polarization as such is not the problem, and those who are concerned about polarization should shift their focus to addressing epistemic issues without abstracting away from content. A paper making this argument is currently undergoing a revise and resubmit.

Reward and Reinforcement Learning

I have also begun a project considering the philosophical significance of reinforcement learning (RL). Artificial RL agents have been able to learn a remarkable range of tasks—most famously, but certainly not exclusively, playing Go at superhuman levels. These exciting advancements have precipitated philosophical speculation. For example, in their seminal RL textbook, Sutton and Barto suggest that 'all that we mean' by goals and purposes can be understood in terms of the maximization of reward as specified by RL. Even more dramatically, David Silver and colleagues argue that 'reward is enough'—that actually *all* cognitive capacities can be understood as subserving the maximization of reward. In an in-progress paper, I scrutinize how we might apply the RL framework to natural agents. The most salient obstacle is that reward is unitary and well-defined for artificial agents because AI researchers specify it by hand. That is, the researchers decide what task they want an RL agent to solve, and so determine what the reward value is for each state of the environment, and build the RL agent so that single-mindedly aims at maximizing long-term reward value. Such clarity of purpose is conspicuously absent for human beings. I consider different ways of overcoming this obstacle—of finding reward in the natural world. A number of psychological candidates can be found to ground the reward role—pleasure, valence, desire, behavior—but each also has drawbacks that limit the comprehensive ambitions some theorists have for applying RL to natural agents. A different possibility I suggest is that RL may be best applied to cognitive subsystems specialized for specific tasks, which alleviates the challenge of identifying a unitary and well-defined reward for the agent as a whole. This doesn't mean that RL is insignificant for the cognitive science of natural agents, as I demonstrate in a commentary applying recent work in reinforcement learning to debates about representational format in perception, which is forthcoming in *Behavioral and Brain Sciences*.