

Research Statement

My primary research questions are: how do we know our own minds, how do others know our minds, and what are these varieties of knowledge good for? My dissertation examines two peculiar epistemic phenomena in connection with these questions. The first is often called *first-person authority*. Roughly, to have first-person authority is to be owed (and typically receive) deference from your listeners when ascribing mental states to yourself. The second phenomenon is the *privileged and peculiar self-knowledge* we tend have of our own mental states. As privileged and peculiar self-knowers, we tend to know our own mental states better than anyone else, and we tend to know them in a way that is unavailable to anyone else. I offer explanations of both phenomena. I then explain how these phenomena intersect in important social contexts. Specifically, I argue that these phenomena jointly contribute to our social agency, which I define as agency that is exercised, for example, in much cooperative group action, interpersonal argumentation, and linguistic interpretation. A more detailed outline of my dissertation (1 page) is attached to the end of my CV.

I have just recently published two papers, which you can view at benwinokur.com/research. I also have two papers under review. One of the papers is currently being revised for resubmission (after receiving a recommendation of acceptance from one referee and a recommendation to revise from a second referee). The other paper has made it past the editor's desk and is currently awaiting reviewer scores.

In the first paper, I respond to recent skepticism about first-person authority (Barz 2018). The skeptic argues that there is either (1) nothing philosophically puzzling about first-person authority, or (2) no such thing as first-person authority. I reply by advancing several refined 'Authority Theses' against the skeptic. The result is that first-person authority is a cluster of interrelated phenomena, each puzzling in ways that warrant philosophical inquiry.

In the second paper, I argue that Donald Davidson's philosophy harbors the resources for an interesting but hitherto underdeveloped transcendental account of *direct* self-knowledge, that is, self-knowledge that is not arrived at by inference, observation, or testimony. On the account, we necessarily have direct self-knowledge insofar as we are capable of interpreting the speech of others in our linguistic communities. I then consider an argument, due to Rockney Jacobsen (2009), that takes Davidson (or the best version of him) to be a deflationist about self-knowledge, one who thinks that the directness of our self-knowledge reduces to the *know how* we possess in using language to self-ascribe mental states. On this view, direct self-knowledge is not propositional knowledge *that* one is in a given state of mind. I deny that Davidson held this view, and I deny that the best version of him would hold it.

I also have several projects on first-person authority and privileged and peculiar self-knowledge still in development. In one paper, I defend a mereological account of privileged self-knowledge of propositional attitudes. On this view, having a certain kind of first-order belief, desire, or intention in the right cognitive conditions partly constitutes having a second-order belief to the effect that one has that belief, desire, or intention, such that no cognitive achievement is required in order to know these attitudes. I defend this account against a battery of criticisms. I also motivate it against a nearby alternative, one that is supposed to better explain the fallibility of our self-knowledge (Parrott 2017).

In a second paper in development, I evaluate different "expressivist" accounts of first-person authority. Expressivist accounts share the idea that listeners are entitled to defer to a speaker's self-ascriptions insofar as those self-ascriptions directly express something about the speaker. But they differ in terms of what expressed feature is relevant. *Neo-expressivists* argue

that the relevant expressed feature is the very mental state self-ascribed (Jacobsen 1996, 1997; Falvey 2000; Finkelstein 2003; Bar-On 2004), whereas *agency-based expressivists* argue that the relevant expressed feature is the agent's cognitive agency with respect to the attitude self-ascribed (Parrott 2015). I defend neo-expressivism against recent criticisms from the agency-based expressivist camp, and I argue against agency-based expressivism directly. Along the way, I draw connections between these expressivist theories and what might reasonably be interpreted as a hybrid expressivist theory (McGeer 2015). I argue that this hybrid theory is not really a hybrid theory so much as a prong of a larger neo-expressivist view. I intend to publish this paper in a special issue of the journal *Philosophies*, tentatively titled "Expression and Self-Knowledge", which I am co-editing with Dorit Bar-On.

A final project within my primary research program is in its earliest stages. This is a monograph titled *The Functions of Self-Knowledge*. This is the first monograph dedicated solely to the question of what functions privileged self-knowledge plays in our psychological economies. The monograph contains six chapters: one introducing questions about the privileged and peculiar character of about self-knowledge, one describing various skeptical accounts of privileged and peculiar self-knowledge, three comprising a critical survey of extant accounts of the functional roles of privileged and peculiar self-knowledge, and one offering my own account—an account that focuses on the indispensability of privileged and peculiar self-knowledge for various forms of social cognition and group action, as originally argued in my dissertation. A detailed chapter outline (approximately 2200 words) is available upon request.

In time, I hope to write papers on the following topics as well:

- 1) The metaphysical differences between 'brute' and 'rational' desires, and the implications of these differences for our self-knowledge of them
- 2) The question of whether self-ascriptions, understood along neo-expressivist lines, can be accommodated by leading theories of testimony
- 3) The question of whether our tendencies to 'confabulate' post-hoc rationalizations for our judgements impedes our first-person authority with respect to those judgements

Moving beyond my primary research program, I am beginning to deepen my interests in social and digital epistemology. I am currently writing a paper about the epistemic injustice one suffers when one's sincere online comments are ignored due to the assumption that they have been posted by a bot. This is increasingly prevalent in North America and the UK in light of recent revelations about the prevalence of Russian bots on platforms like Reddit and Twitter. Of course, because such bots are real, many "accusations of bothood" are true. But some are false. The false accusations are interesting, I argue, because they generate epistemic injustices that look a lot like "testimonial injustices", except that they are not rooted in "identity-prejudicial credibility deficits" (Fricker 2007). After all, as testimonial injustice is typically understood, the speaker's testimony is not taken *sufficiently seriously* owing to the hearer's prejudices against the speaker due to the speaker's sex, race, class, etc. However, when someone assumes that a post has been made by a bot, it is being assumed that *no testimony has really been issued*, since bots are not genuine speakers. I explore some consequences of this fact for thinking about how to combat epistemic injustices produced by accusations of bothood.

References:

- Bar-On, D. 2004. *Speaking My Mind: Expression and Self-Knowledge*. New York: Oxford University Press.
- Barz, W. 2018. Is There Anything to the Authority Thesis? *Journal of Philosophical Research*, 43. doi: 10.5840/jpr2018712122
- Byrne, A. 2005. Introspection. *Philosophical Topics*, 33(1), 79-104.
- Byrne, A. 2011. “Knowing What I Want” In J. Liu & J. Perry (Eds.), *Consciousness and the Self: New Essays*. Cambridge, UK: Cambridge University Press.
- Byrne, A. 2018. *Transparency and Self-Knowledge*. Oxford University Press.
- Falvey, K. 2000. The Basis of First-Person Authority. *Philosophical Topics*, 28(2); 69-99.
- Finkelstein, D. 2003. *Expression and the Inner*. Harvard University Press.
- Fricke, M. 2007. *Epistemic Injustice: Power & The Ethics of Knowing*. New York: Oxford University Press.
- Jacobsen, R. 1996. Wittgenstein on Self-Knowledge and Self-Expression. *The Philosophical Quarterly*, 46(182), 12-30.
- Jacobsen, R. 1997. Semantic Character and Expressive Content. *Philosophical Papers*, 26(2), 129-146.
- McGeer, V. 2015. Mind-Making Practices: the Social Infrastructure of Self-Knowing Agency and Responsibility. *Philosophical Explorations*, 18(2): 259-281.
- Parrott, M. 2015. Expressing First-Person Authority. *Philosophical Studies*, 172(8), 2215-2237.
- Parrott, M. 2017. Self-Blindness and Self-Knowledge. *Philosophers' Imprint*, 17(16), 1-22.