

## Research Statement

How do you know your own mind, how do others know your mind, and what good is such knowledge? My doctoral thesis examines two peculiar epistemic phenomena in connection with these questions. The first is our *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 nobody else can know them. I offer explanations of both phenomena, and I explain how these phenomena intersect in important psychological contexts. Specifically, I argue that these phenomena jointly contribute to our social agency, which I define as agency that is exercised in much of our cooperative group action, interpersonal argumentation, and linguistic interpretation. A more detailed outline of my doctoral thesis (1 page) is attached to the end of my CV.

I have just recently published two papers, which you can learn more about at [benwinokur.com/research](http://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 under review, 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, in fact, a cluster of interrelated phenomena, each puzzling in ways that warrant philosophical inquiry.

In the second paper under review, I argue that Donald Davidson's philosophical corpus harbors the resources for an interesting but hitherto underdeveloped transcendental account of why we must have a peculiar form of access to our own minds. On the account, we necessarily have peculiar 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 have a deflationary conception of self-knowledge, according to which our self-knowledge reduces to the *know how* we possess in being able to self-ascribe our mental states (as opposed to understanding it as propositional knowledge). 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 both first-person authority and privileged and peculiar self-knowledge still in development. I defend a "constitutivist" account of our self-knowledge of our propositional attitudes. On this account, a first-order belief, desire, or intention, held in the right cognitive conditions, is itself a *part* of one's second-order belief to the effect that one has it, rather than an ontologically independent mental state that one must *detect* in order to possess self-knowledge. I defend my constitutivist account against a battery of criticisms, such as the criticism that it violates the independently plausible thesis that knowledge always involves a cognitive achievement, and the criticism that it is susceptible to a vicious regress (Greene 2003). Finally, I defend my account 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. First-person authority is the idea that an agent's self-ascriptions of her current mental states ought to be presumed true, at least typically, even without the agent's presenting any positive evidence for their truth (or even *being able to present it*). Expressivist explanations

of first-person authority 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 also evaluate what might be interpreted as a hybrid expressivist theory (McGeer 2015), and I argue that neo-expressivism can accommodate it as a special case. I eventually 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 roles privileged and peculiar self-knowledge plays in our psychological economies. The monograph contains six chapters: one introducing key intuitions about the privileged and peculiar character of about self-knowledge, one describing various skeptical responses to these intuitions, 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 doctoral thesis. A more detailed précis (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 how we acquire self-knowledge of them
- 2) The differences, if there be any, between self-ascriptive speech acts that *express* one's mental states and self-ascriptive speech acts that *testify* as to one's mental states
- 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 have been deepening my familiarity with current topics 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 the hearer's "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
- 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.
- Fricker, M. 2007. *Epistemic Injustice: Power & The Ethics of Knowing*. New York: Oxford University Press.
- Greene, R. 2003. Constitutive Theories of Self-Knowledge and the Regress Problem. *Philosophical Papers*, 32(2): 141-148.
- 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.