

Research Statement

How do you know your own mind, how do others know your mind, and what is the importance of this knowledge? Most of my research investigates two epistemologically puzzling phenomena in connection with these questions. The first phenomenon 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. Roughly, in having privileged self-knowledge, one knows one's own mental states better than anyone else, and in having peculiar self-knowledge one knows one's mental states in a way that is only available to oneself. In my work, I seek to understand the origins, nature, and psychological import of these phenomena, and to better understand the ways in which they intersect.

You can learn more about my published works at benwinokur.com/research. Here, I will describe my not-yet-published projects. I begin with two papers currently under review. The first has received a conditional acceptance, pending minor revisions. In it, I respond to recent skepticism about first-person authority. 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 respond to this dilemma by proposing several refined specifications of first-person authority.

In a second paper under review, I argue that Donald Davidson's philosophical corpus contains the resources for an interesting but hitherto underdeveloped transcendental account of peculiar self-knowledge. On this account, we necessarily have peculiar self-knowledge insofar as we are capable of interpreting the speech and thoughts of others. This has the interesting consequence of showing that one of the chief functional roles of peculiar *self-knowledge* in our psychological lives has to do with our ability to know *other* minds. Along the way, I critically address a reading of Davidson according to which he was a kind of deflationist or quietist about self-knowledge, i.e., as someone who thought that privilege self-knowledge was either reducible to a capacity for first-person authoritative *speech*, or as someone who thought that peculiar self-knowledge could not be explained at all.

In a paper close to completion, I evaluate different "expressivist" accounts of first-person authority. Expressivist explanations of first-person authority share the idea that listeners are entitled to (and do) defer to speakers' self-ascriptions insofar as those self-ascriptions directly express something about the speaker, but they differ on the question of what expressed items are explanatorily relevant. *Neo-expressivists* argue that the relevant expressed items are the very mental states self-ascribed, whereas *agency-based expressivists* argue that the relevant expressed item is the agent's *capacity* to determine her mental states in the manner of a cognitive agent. I defend neo-expressivism against criticisms from the agency-based expressivist camp, after which I argue against agency-based expressivism directly. I will soon publish this paper in a special issue of the journal *Philosophies*, tentatively titled "Expression and Self-Knowledge", that I am co-editing with Dr. Dorit Bar-On.

Another paper, currently in development, identifies an as-yet-unappreciated aspect of our first-person authority, namely, the distinctively *easy warrant* that others often have to defer to what we say about our own minds. Upon arguing that this is indeed an aspect of first-person authority, I go on to offer two interesting consequences: (1) first-person authority likely ought *not* to be explained in terms of the privileged and peculiar self-knowledge that speakers have of their mental states, and (2) a certain conception of testimonial justification—one that appeals to a capacity for *virtuous hearing* (as in Miranda Fricker's *Epistemic Injustice*)—gains plausibility.

Within my primary research program, a final work in progress is in its earlier stages. This is a monograph titled *The Functions of Self-Knowledge*. This is the first monograph dedicated solely to the question of what functional roles privileged and peculiar self-knowledge plays in our psychological economies. The monograph will contain six chapters: one introducing key intuitions about the privilege and peculiarity of much of our 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 of the indispensability of privileged and peculiar self-knowledge for various *social-epistemic* ends, as originally argued in my doctoral thesis. A more detailed précis of this monograph (approximately 2200 words) is available upon request.

Here are some topics that I hope to turn to as I wrap up some of the above projects:

- 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) Whether there are multiple ways in which one can adopt a first-person perspective on one and the same mental state
- 4) Whether certain recent accounts of the social functions of reasoning have implications for the indispensability of self-knowledge and other-knowledge

Moving beyond my primary research program, I have recently been deepening my interests in social and digital epistemology. In a paper currently under review, I consider cases where one’s sincere and competent online comments are ignored due to the assumption that they have been posted by a bot. Such cases are increasingly prevalent in North America and the UK in light of recent revelations about the prevalence of (primarily Russian) bots on platforms like Reddit and Twitter. Of course, because many such bots are real, “bot judgements” are often true. But some are false. The false ones are interesting, I argue, because they seem to generate epistemic injustices that look a lot like ‘testimonial injustices’, except that they are not rooted in the hearer’s “identity prejudices”. As testimonial injustice is typically understood, a speaker’s testimony is not taken *sufficiently seriously*, owing to the hearer’s prejudices against speakers of a given sex, class, ethnicity, and so on. However, when someone judges that a post has been authored by a bot, it is being judged 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 false bot judgements.

In social and digital epistemology, I am also beginning to focus on questions like:

- 1) In what ways ‘deepfake’ technologies might spur *positive* epistemic changes in our societies, despite the many obviously bad epistemic consequences that these technologies threaten
- 2) Whether testimonial injustice always requires hearers to behave epistemically or morally irresponsibly toward speakers

I encourage you to contact me if you are interested in hearing more about any of my works!