

A Virtue Epistemology of Trust

research project, 2020-2023

PI: J. Adam Carter

Co-Is: Christoph Kelp and Mona Simion

Funded by: The Leverhulme Trust

One of the most serious challenges faced by philosophers of trust is to understand why, and under what circumstances, we should trust as opposed to distrust others and what they tell us (e.g., Hawley 2014; Faulkner 2011; Nickel 2001). Even though philosophical theories of trust have offered insights into what trust is (e.g., Baier 1986, Jones 1996; McMyler 2011), they have yet to tell us what qualities make someone a good or bad truster, and how they do so.

This project introduces virtue epistemology (e.g., Sosa 2009; 2015; Greco 2010; Carter 2019; Kelp 2018c) to address this issue for the first time. It offers a novel method for theorising about what dispositions trusting well requires, and it uses this method to explain why certain forms of skilled trusting are more valuable than others.

The project is also designed to break new ground in debates about trustworthiness by showing that skills needed to be a good truster, as well as to be reliably trustworthy, are importantly related. Further, the project connects these results more widely to debates in social epistemology by showing how trusting well provides a constraint on three key aspects of social life: assertion, action, and practical reasoning.

An important payoff is an understanding of how skilled trusting can help build and sustain more resilient trust networks. It is urgent that we gain such answers and insights: according to the 2018 Edelman Trust Barometer, UK public trust in social media and online news has plummeted to below 25%, and trust in government is at a low 36%. This present crisis in trust of corresponds with a related crisis of distrust, in that the dissemination and uptake of fake news, particularly on social media, have risen dramatically the past few years (Levinson 2017; Barclay 2018; Lynch 2019). Public policy regulations (e.g., Klein & Wueller 2017) can at best treat the symptoms of these problems in the absence of a deeper understanding of their sources in our own agency (Carter 2019).

Questions that the project will seek to answer:

Q1: What kind of traits make a person a good truster, and how do they do so?

Q2: What is the relationship between being a good truster, and having the kind of traits in virtue of which one is trustworthy?

Q3: How can thinking about what's involved in trusting well help us to build more resilient and healthy trust networks?

Getting clear answers to these questions would constitute a much-needed breakthrough in the philosophy of trust, one that is analogous to the breakthrough that the virtue-theoretic framework has already provided epistemologists for thinking about knowledge. And, practically speaking, rigorous and principled answers to Q1- Q3 will go a long way to helping us to isolate key sources of trust's erosion that lie within our own agency and control.

Objectives

The core hypothesis to be tested is that insights from virtue epistemology can transform our understanding of what it is to trust well, and in a way that will help us make needed traction on each of Q1-Q3, questions which are philosophically important and—given the widespread erosion of trust—politically and socially timely. The proposed project offers—in a way that is informed by cutting-edge work in virtue epistemology and the philosophy of trust—a novel virtue-theoretic account of: 1. the nature and value of trust; (in connection with Q1) 2. its structural relationship to trustworthiness; (in connection with Q2) 3. its normative connections with the social-epistemic practices of assertion, action and practical reason (in connection with Q3).

In a bit more detail, we investigate for the first time how trusting, like believing, corresponds with certain kinds of trust-specific dispositions to manage interpersonal risks in ways that can be more or less reliable. Focusing on what these dispositions look like (both generally, and when they manifest in successful trust) will tell us a lot about what specific shapes trusting well will take, why some are more valuable than others (epistemically and morally), and how skilled trusting connects to other things that are vital to successful social and political life, such as being a good asserter, agent, and deliberator.

We take as a starting point in our attempt to understand trust a key analogy between trust and belief that has not been suitably developed, and which concerns their respective constitutive aims. A widely-shared view in epistemology, and in the literature on norms of belief (e.g., Chan 2013), is that belief aims at truth in the sense that a belief that is not true is, as such, in some sense defective—viz., it misses its mark (e.g., Shah & Velleman 2005; Wedgwood 2002; cf., Simion et al. 2016). Does trust constitutively aim at anything? A *prima facie* case for thinking 'yes' is that trust has clear fulfillment conditions in paradigmatic cases of interpersonal trust: X's trusting Y to do some action is fulfilled only if Y has taken care of things as X has entrusted them (Carter 2019). There is much discussion about what it would be for the trustee to take care of things as entrusted—viz., what additional attitudes and/or beliefs this may involve, including attitudes and beliefs about the attitudes (e.g., goodwill) and beliefs of the trustee (e.g., Baier 1986; Jones 1996; cf., Hawley 2014). But however these details are filled out, it remains that when one trusts another to take care of things as entrusted, there is a clear sense in which the trust succeeds only when the trustee actually does take care of things as they are entrusted, perhaps *compresently* with certain attitudes or beliefs. Accordingly, just as belief is successful only if it is true (e.g., when mind fits world), trust is successful only if it is fulfilled (e.g., when that trust is not betrayed).

This initial analogy between belief and trust lends itself to interesting further

development through the application of performance-theoretic virtue epistemology (e.g., Sosa 2015), which maps out the normative structure of performance-types with aims that are internal to that type of performance. For instance, following Sosa, we can think of the normative structure of a belief along three dimensions: success, skill, and aptness. A belief, successful if true, might nonetheless be successful through sheer luck. It is skillful only if it issues from dispositions that would ordinarily generate true beliefs when the subject is in proper shape and properly situated. But even successful and skillful beliefs can fall short of a higher grade of assessment—aptness—if the success does not manifest the skill exhibited, but is instead due to luck. The idea of apt trust takes us beyond previous assessments of trust’s nature and value. Even more, there is scope to investigate within this normative framework how the more sophisticated aim of apt trust (and not merely the aim of successful trust) might itself be attained in ways that manifest our higher-order trust-relevant abilities, e.g., abilities to monitor for risk of inapt trusting (cf., Sosa 2015; Carter 2019; Kelp et al. 2017), and not merely risk of betrayed trust. Along with providing an entirely new virtue-theoretic account of trust, we will show how this account naturally complements a virtue-theoretic account of trustworthiness (cf., Potter 2002), and also how it comports well with three key aspects of social-epistemic practice: assertion, action, and practical reason. In particular, we explore how our core notion of apt trust plausibly functions as a norm governing certain kinds of assertions and actions internal to the practice of giving and receiving trust, and we articulate the conditions under which violations of this norm, as we characterize it, may or may not be blameless (e.g., Kelp and Simion 2017; Kelp 2019; Williamson Forthcoming).

The overarching project will involve three phases—each of which corresponds with one of the three central project objectives—and it will be led throughout by the PI and Co-Is, with input from an expert steering committee. The project postdoctoral fellow and Ph.D student will be centrally involved in all aspects of the project. A summary plan for each project phase—including key research outputs, events, and other activities—is as follows:

Phase 1. Performance Epistemology in the Philosophy of Trust (1 January 2020 - 1 January 2021) Phase 1 will lay the groundwork for an application of performance-theoretic virtue epistemology in the philosophy of trust in three stages: (i) first, we map the relevant terrain in performance epistemology and the philosophy of trust by surveying the recent literature in each area with the aim of taxonomizing relationships between key concepts; (ii) second, we develop a version of the performance-theoretic framework that we will defend as most suitable for theorising about trust; and (iii) we outline key desiderata for applying the account successfully. Regarding (i): An area we will be exploring for the first time in Phase 1 is how the three-part performancenormative structure that has been well-explored with belief could apply, *mutatis mutandis*, to trust, and how an account of skilled and apt trusting might best be developed with reference to the notion of successful trust in a way that is broadly analogous to how an account of skilled and apt believing has been developed successfully with reference to successful (true) belief. By carefully investigating both the trust and virtue epistemology literature, we will

see how far we can push these performance-theoretic analogies in an effort to connect the two debates in ways they've not been thus far. Moreover, in relation to (ii) and (iii), an account of skilled trusting will require the identification of dispositions that will help us better manage our interpersonal risks under certain conditions. But which conditions, exactly, are the relevant ones for the purpose of assessing our reliability in doing so? Do they include, or exclude, conditions where, for example, the risk to the truster is excessively high and gains of betrayal are enormous, and/or where the level of effort or skill that would be required by the trustee to take care of things as entrusted is unusually high? Investigating these boundaries carefully, and how they relate to skilled trusting, bears importantly on how skilled trusting should be modeled within a virtue-theoretic framework. Furthermore, it will be important to have a clear sense of what additional kinds of desiderata we should expect a successful application of the performance-theoretic model to trust to satisfy. Among other things, it should be compatible with empirical work on the moral psychology of trust, and it should be capable of explaining sources of misplaced trust.

Phase 2. Trust and Trustworthiness (1 January 2021 - 1 January 2022) Phase 2 has two central objectives: (i) the development and defence of a virtue-theoretic account of trust; and (ii) a virtue-theoretic account of trustworthiness, where the latter involves a way of actively and positively engaging with the fact of the trustee's dependency, made possible by our capacity to recognize such dependency and take it into account in our deliberation (Jones, 2012). Whereas extant literature often fails to show how these two notions are connected, we will defend a view of trust on which the two are essentially interdefinable; this project draws from the resources of the network model of analysis pioneered in recent epistemology by Kelp (2018a). Once our account is developed and refined, we show how the interdefinability of trust and trustworthiness affords us the resources to overcome traditional problems in the epistemology of trust, including the problem of monitoring. Page 6 of 21

Executing task (i) will involve further development of the model of trust proposed in Phase 1 of the project. In particular, we will show how a complete account of trust and trusting well needs to make room for two different 'levels' of trust-related skills. There are those outlined in Phase 1, which are important for apt trust, and which demand reliability in trusting successfully. But there are also reflective (i.e., second-order) skills that are important to reliably assessing the risk of inapt trust. These reflective skills require not only reliable (first-order) risk assessment, but also the reliable assessment of our own capacities to assess these risks. Regarding task (ii): Well-functioning trust networks require trustworthiness in addition to skilled trusting. To a first approximation, a trustee B is trustworthy with respect to a truster, A, in a domain of interaction D, if and only if she is competent with respect to that domain, and she would take the fact that A is counting on her, were A to do so in this domain, to be a compelling reason for acting as counted on (Jones 2012). One standard way to theorise about trustworthiness would have us break these two conditions down into further sub-conditions. This method is called, following Strawson (1992), the dismantling model of analysis, a method whose shortcomings have been increasingly highlighted (e.g., Williamson 2000). Drawing from work by Kelp (2018a) on network models of analyses, we will instead locate trustworthiness,

along with trust, as key nodes within a wider social-epistemic network, by showing how an understanding each involves at least some understanding of the other.

Phase 3. A Virtue Epistemology of Trust: Normativity, Value, and Social Epistemology (1 January 2022 - 1 January 2023) Phase 3 has two central objectives; first, an axiology of trust-related achievements will be articulated—i.e., what makes them distinctly valuable, and how this motivates a particular conception of flourishing within a trust network. Secondly, and drawing from recent work on epistemic axiology by the PI and Co-Is (Carter 2017; Carter et al. 2015; Kelp 2016; Simion 2016b, 2016a, 2018), we develop an account of the normativity of trusting according to which the notion of apt trust provides a normative constraint or rule governing certain kinds of actions and mental states that feature prominently in well-functioning trust networks. In particular, we defend the view that apt trust is the norm of trusting, as well as the norm of certain kinds of assertions, actions and reasoning patterns that constitute moves within a trust network. Regarding (i): Practices are organised around values; the practice of inquiry, for example, is plausibly organised around the value of knowledge (Kelp 2018c; cf., Craig 1990) or perhaps understanding (Kvanvig 2013). Norms or rules internal to a practice are connected to the values around which the practice is organised, and such norms plausibly sustain a practice when the value that is achieved by following such rules suffices to explain why agents continue to follow them (e.g., Turri 2017). A clear view of the values governing the practice of giving and receiving trust will be important to characterising the norms internal to it, and such a view should be compatible with the thought that the value of following them would be important to sustaining the practice. In Phase 3 of the project, our initial aim is to show how certain kinds of achievements in trusting, successes that manifest ability (e.g., Sosa 2009; Greco 2010; Carter et al. 2015; Bradford 2015), can play this kind of axiological role, and how practice-sustaining norms can be articulated in terms of them. Regarding (ii): We develop this idea in more sophistication with reference to three specific kinds of socialepistemic norms that are central topics within social epistemology: norms governing assertion, action, and Page 7 of 21 practical reasoning. In particular, and drawing from the project team’s extensive work on these topics, we will articulate how it is that trusting and trustworthiness, and the thought, action, and communication that matters for both, are normatively constrained.

References

- Alfano, M. (2016). 'The Topology of Communities of Trust', *Russian Sociological Review*, 15(4), 30–56.
- Baier, A. (1986). 'Trust and Antitrust', *Ethics*, 96(2), 231–260.
- Baker, J. (1987). 'Trust and Rationality', *Pacific Philosophical Quarterly* 68 (1), 1–13.
- Brady, M., & D. Pritchard. (2006). 'Epistemic Virtues and Virtue Epistemology', *Philosophical Studies* 130 (1), 1–8.
- Carter, J. A. (2016). 'Sosa on Knowledge, Judgment and Guessing', *Synthese*, 1–20. doi:10.1007/s11229-016-1181-2

Carter, J. A. (2017). 'Assertion, Uniqueness and Epistemic Hypocrisy'. *Synthese* 194(5), 1463-1476.

Carter, J. A. (2019). 'On Behalf of a Bi-Level Account of Trust', *Philosophical Studies*, 1-24. doi:10.1007/s11098-019-01311-2

Carter, J. A., B. Jarvis & K. Rubin. (2015). 'Varieties of Cognitive Achievement', *Philosophical Studies* 172 (6), 181–199.

Chan, T. (Ed.). (2013). *The Aim of Belief*. Oxford: Oxford University Press. • Faulkner, P. (2011). *Knowledge on Trust*. Oxford: Oxford University Press.

Fernandez, M. A. (Ed.). (2016). *Performance Epistemology*. Oxford: Oxford University Press.

Fricker, E. (1995). 'Critical Notice: Telling and Trusting—Reductionism and Anti-Reductionism in the Epistemology of Testimony', *Mind* 104(414), 393–411.

Frost-Arnold, K. (2014). 'The Cognitive Attitude of Rational Trust', *Synthese* 191(9), 1957–1974.

Gambetta, D. (1988). 'Can We Trust Trust?' In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations* (pp. 213–237). Oxford: Blackwell.

Goldman, A.I. (2001). 'Experts: Which Ones Should You Trust?', *Philosophy and Phenomenological Research*, 63(1), 85–110.

Greco, J. (2009). 'Knowledge as Success From Ability', *Philosophical Studies* 142, 17–26. • Greco, J. (2010). *Achieving Knowledge*. Cambridge: Cambridge University Press.

Greco, J. (2012). 'A (Different) Virtue Epistemology'. *Philosophy and Phenomenological Research* 85, 1–26.

Hardin, R. (2002). *Trust and Trustworthiness*. New York City: Russell Sage Foundation.

Harding, M. (2011). 'Responding to Trust', *Ratio Juris*, 24(1), 75–87. • Hawley, K. (2014). 'Trust, Distrust and Commitment', *Noûs*, 48(1), 1-20. Page 9 of 21

Hieronymi, P. (2008). 'The reasons of Trust', *Australasian Journal of Philosophy* 86(2), 213–236. • Jones, K. (2012). 'Trustworthiness', *Ethics* 123(1), 61–85.

Jones, K. (1996). 'Trust as an Affective Attitude', *Ethics* 107(1), 4–25. • Kelp, C.. (2018a). 'Inquiry and the Nature of Knowledge' (ms.)

Kelp, C.. (Ed.). (2018b). *The Epistemology of Ernest Sosa*. Special issue of *Synthese*.

Kelp, C. (2018c). *Good Thinking: A Knowledge First Virtue Epistemology*. London: Routledge. • Kelp, C. (2019). 'The Knowledge Norm of Blaming.' *Analysis*.

Kelp, C., et al. (2017). 'Hoops and Barns: A New Dilemma for Sosa', *Synthese*. doi:10.1007/s11229- 017-1461-5.

Kelp, C., & Simion, M. (2017). 'Criticism and Blame in Action and Assertion', *Journal of Philosophy* 114(2), 76-93.

Klein, D. & Wueller, J. (2017). 'Fake News: A Legal Perspective'. SSRN Scholarly Paper ID 2958790. Rochester, NY: Social Science ResearchNetwork. <https://papers.ssrn.com/abstract=2958790>.

Kvanvig, J. (2013). 'Curiosity and a Response-Dependent Account of the Value of Understanding', In T. Henning and D. Schweikard (eds.), *Knowledge, Virtue and Action*. Boston: Routledge.

Levinson, P. (2017). *Fake News in Real Context*. New York: Connected Editions, 2017

Livingstone, S. et al. (2019) 'Tackling the Information Crisis: A Policy Framework for Media System Resilience'. • McLeod, C. (2015). 'Trust', In E. N. Zalta (Ed.), *The Stanford Sncyclopedia of Philosophy*. Palo Alto, CA: Stanford University.

McMyler, B. (2011). *Testimony, Trust, and Authority*. New York: Oxford University Press.

• Möllering, G. (2006). *Trust: Reason, Routine, Reflexivity*. Elsevier

O’Neil, C. (2012). 'Lying, Trust, and Gratitude', *Philosophy and Public Affairs*, 40(4), 301–333

• Potter, N. (2002). *How Can I Be Trusted? A Virtue Theory of Trustworthiness*. Rowman & Littlefield.

Shah, N., & Velleman, J. D. (2005). 'Doxastic Deliberation', *The Philosophical Review* 114(4), 497-534.

Simion, M. (2016a). 'Assertion: Just One Way to Take It Back', *Logos & Episteme* 7(3), 385–391.

Simion, M. (2016b). 'Assertion: Knowledge Is Enough', *Synthese* 193(10), 3041–3056.

Simion, M. (2018). 'No Epistemic Norm for Action', *American Philosophical Quarterly* 55(3), 231–238.

Simion, M., Kelp, C., & Ghijsen, H. (2016). 'Norms of Belief', *Philosophical issues* 26(1), 374-392.

Sosa, E. (1991). *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press.

Sosa, E. (2009). *A Virtue Epistemology: Apt Belief and Reflective Knowledge (Vol. I)*. Oxford: Oxford University Press.

Sosa, E. (2010). 'How Competence Matters in Epistemology', *Philosophical Perspectives* 24(1), 465– 475.

Sosa, E. (2011). *Knowing Full Well*. Princeton: Princeton University Press.

Sosa, E. (2015). *Judgment and Agency*. Oxford: Oxford University Press.

Strawson, P. F. (1992). *Analysis and Metaphysics: An Introduction to Philosophy*. Oxford: Oxford University Press.

Turri, J. (2013). *Virtuous Thoughts: The Philosophy of Ernest Sosa*. Dordrecht: Springer.

Turri, J. (2017). 'Sustaining rules: a model and application', In J. A. Carter, E. C. Gordon, & B. Jarvis (Eds.), *Knowledge First*. Oxford: Oxford University Press.

Sosa, E. (2017). *Epistemology*. Princeton: Princeton University Press.

Wedgwood, R. (2002). 'The Aim of Belief', *Philosophical Perspectives* 16, 267-297.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, T. (Forthcoming). 'Justifications, Excuses, and Skeptical Scenarios', (Ed.) J. Dutant & F. and F. Dorsch. *The New Evil Demon*. Oxford: Oxford University Press.

Zagzebski, L. T. (1996). *Virtues of the Mind*. Cambridge: Cambridge University Press.

Zagzebski, L. T., (2012). *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. New York: Oxford University Press.