

The Material Theory of Values in Science

Kelli Barr (Pittsburgh, Pennsylvania, United States)

Abstract

How are we to understand situations where science fails on its own terms? Scientists have blamed perverse incentives for systematic epistemic failures like non-replicability and publication bias, but the exact relationship remains an open question. Let's assume they are right to blame the (social) system. This paper presents a novel framework for understanding how features of the social organization of science are implicated in collective epistemic failures: the material theory of values in science (MTV). This project is inspired by and follows in the tradition of feminist philosophers of science who have called attention to the need for explanations of systemic, specifically antifeminist, biases in science and for embodied models of scientists as epistemic agents. In the first part, I discuss the replication crisis as involving a particular type of collective action problem: a no-win standoff. The next part introduces the MTV and the explanation it supplies for this phenomenon. In the third section, the MTV is compared to several alternative explanatory strategies, including from the contemporary literature values in science and social epistemology, specifically agent-based computational models. I argue for why my approach is preferable and describe an important revision it entails for the general Mertonian sociological picture invoked in discussions of incentives in science. By way of conclusion, I summarize the philosophical task of the MTV, which is to take seriously the sociological and material conditions constituting scientists' epistemic agency.

Since revelations of the replication/reproducibility crisis in many research specialties, scientists are in a situation characterized not only by publish-or-perish, but also by the possibility that their findings might not be valid, secure, or even true despite their best efforts. Before this, publish-or-perish already called to mind the classic movie trope of a standoff: like adversaries circling each other with weapons drawn, scientists competed for who would be the first to publish the next discovery by continuously escalating the pace and output of publications. This was a game one could win, in principle, by beating others to the publication punch. However, the replication crisis has turned this situation into one where there is no path to victory for individual researchers, since systemic biases and failures manifest in aggregate activity (OSC 2015; Nosek et al. 2022). It is a no-win standoff; it can only be resolved by mutual destruction (i.e., dropping out of the profession) or mutually withdrawing. One cannot unilaterally withdraw, for example, by running a replication or reproduction study to further validate one's own or others' existing findings without taking oneself out of the running (at least temporarily) for being the first to publish a discovery. But convincing others to stand down collectively is terribly difficult. Most who remain in the profession remain in the standoff.

Scientists have tried to bargain with one another for mutual withdrawal. Some have tried to target incentives directly, for example by pursuing special funding for large-scale replication/reproduction projects or starting special venues for publishing such projects, in the hopes of shifting collective activity toward more validation of findings (see Korbmacher 2023 for a more comprehensive review of such efforts). Others have thrown their efforts into developing Metascience as a research specialty for scientific work about science itself, on the hopes of redeeming the enterprise with its own best tools. One persistent problem has been how to understand claims that systemic epistemic failures and biases are due, in part, to features of the social organization of science. Perverse incentives are consistently cited in discussions of the replication crisis. To take one hypothesis as an example: publish-or-perish might drive researchers to cut corners using questionable research practices. Many scientists seem convinced that there is some connection, but the exact relationship remains an open question.

Let's assume scientists are right that (at least some) epistemic failures are due to social factors. What can we learn from this? First, we can learn that the situation is not best explained by what scientists believe, or prefer, or know, or have attitudes about. A no-win standoff is a standoff precisely in virtue of the weapons each party points at one another, fingers poised ready on the triggers. But incentives aren't guns to peoples' heads... right? When it comes to the replication crisis, scientists talk as though they are compulsory. This is more than just a way of speaking. There are material consequences to withdrawing from the standoff. At stake is continued access to the resources necessary to keep doing research at all, such as access to funding, labor, and equipment – meaning, the possibility of continuing to have a career. There are many ways to think about materiality, notably stemming from Marxist intellectual traditions. My concern here is primarily with the means of scientific production listed above, but also secondarily with matters of socially situated agency and embodiment (Haraway 1988; Rouse 2002), with social ontology (cf. Haslanger 2022), and with explanations of systemic biases in critiques of science (Wylie 1996; Lloyd 2006). Second, we can learn about a broader field of values that structure scientific work by paying attention to the material dimensions of the social organization of science.

Statement of the Theory

The material theory of values in science (MTV) holds that there is a class of values that shape scientific practices independently of individual scientists' beliefs, preferences, knowledge, attitudes, or other cognitive phenomena. I call these *background values*, which are constituent parts of the *institutions* that structure scientific practices and are evident in scientists' interactions with their *material conditions*: the organization, allocation, and distribution of money, labor, and equipment. Scientists' actions, then, aren't explained by cognitive phenomena alone, or perhaps even primarily, but rather by their enacting the institutions that structure their work and the background values made material by those institutions.

MTV has two core tenets. First: values are materialized in institutions. To enact an institution is to enact a set of values. This has the consequence that values are not the kinds of things that one holds, like attitudes, or that one takes for granted, like background assumptions. Like the institutions which give them concrete expression, they are shared, collective things. By institutions, I mean those stable but flexible patterns of conventional conduct that structure individual and collective social behavior. This sense of 'institution' is inspired by the Deweyan pragmatist school of sociology based at the University of Chicago, which was most active in the middle and late part of the 20th century (see Hughes 1942; Gerson, unpublished manuscript). My use is more similar to what others characterize as 'social practices' (e.g., Rouse 2003) than to typical conceptions of institutions, especially among philosophers (see, for example, Turner 1994; Guala 2016). Hamburgers, putting on socks and shoes, the time of day when espresso is drunk, the particular layout of a biological laboratory's storage cabinets, and funding schemes; all these count as institutions and exist along interrelated scales of formality, longevity, complexity, and reach. The claim that values are constituents of institutions, then, entails that we have values as a result of the historical, ongoing societal organization of collective life, including the collective life of scientific communities. Values do not exist as free-floating, psychological, ethical/moral, or existential parameters. They are not intrinsic to human cognition or agency. To say that you value something is to claim membership in a community and to be (or to have been) shaped by that membership.

The second tenet of the MTV is that values have a material, rather than formal or psychological, relation to institutions. This contrasts with long running discussions in philosophy of science about the possible formal relations between values and, for example, theory choice or hypothesis acceptance (cf. Rudner 1953). Contemporary discussions of values in science motivated by the argument from inductive risk tend to treat values as decision criteria or parameters that stand in logical, justificatory, or causal relationships to individual decisions made by individual scientists. Values are taken to exist in the minds, heads, and decisions of individuals; for such cognitivist theories, the instantiation of values is logical or psychological. By contrast, I say that values exist in institutions; for material theories, the instantiation of values is sociological and intersubjective. This addresses a persistent problem in the values and science literature: how to understand the nature of the values we think are operative in science (Elliott and Korf 2024). Ward (2021) proposed an initial taxonomy: values can be reasons (for example, for actions or beliefs), or they can be causes (for example, as motivations). To this, the MTV adds that values can be constituent features, as in part-whole compositional relations, which locates the normativity of values in the institutions of which they are constituent parts rather than in individual psychological dynamics, decision-making, or utility maximizing.

Let's return briefly to the situation described at the outset, with the MTV in hand. Scientists identify publish-or-perish as a contributor to systemic epistemic biases like those toward publishing novel and positive findings. The idea is that incentives prioritizing novelty direct scientists to put their limited time and resources toward new discoveries instead of validating existing findings. According to our no-win standoff model, no one is willing to incur the risks of switching tracks from discovery to validation unless or until others do the same. Exactly how and why is this the case? The MTV directs us to examine the institutions implicated in publish-or-perish. For which scientific practices is productivity, in terms of high publication output, a constitutive value and how does its value relate to scientists' interactions with material conditions in which they work? We are led to observations of this kind: productivity is a constitutive value of such institutionalized practices as decisions about funding allocation, hiring, promotion, and tenure. That is, publications regulate access to the material goods and resources without which one cannot

continue to do research. One needs publications to get jobs, research funding, research labor, and other resources, like access to research libraries and laboratories, or other data collection equipment. To the extent that novel discoveries are a priority, in terms of it being much more difficult to get validation projects published, and to the extent that publications are required to continue on with a research career, publish-or-perish is a driver of publication bias. In this sense, producing epistemically sub-optimal outcomes is constitutive of the (current) material and social conditions of science, without anyone involved wanting this to be the case, or even trying to make it so.

Why Blame Institutions and Not Individuals?

One significant advantage the MTV has over alternatives lies in highlighting the dual economies in which publications are the coin of the realm. On the one hand, publications are symbolic currency in the reputational economy of credit (e.g., for discoveries) and prestige (e.g., honorary doctorates, exclusive society memberships). On the other hand, they are currency in the material economy that allocates financial, labor, and equipment resources. In this sense, the MTV motivates a revision to a common sociological picture in the background of discussions about the structure of scientific communities. Sociologists of science have long recognized the importance of credit allocation in incentivizing researchers to publish. Merton's (1957) priority rule formalized this relationship, and some philosophers of science have since championed it as key to achieving an efficient distribution of labor between scientists and between research programs (Kitcher 1999, Strevens 2011). Recent discussions are less sanguine about the efficiency of such a system (Romero 2017, Heesen 2018). But they do not dispute the basic features of the model: scientists pursue projects likely to garner recognition from their peers, in which case maximizing productivity is an advantageous strategy for maximizing recognition.

As the MTV shows, however, the focus on *what* scientists choose to publish belies institutional realities. Scientists cannot make a career from publishing replications and/or reproductions of others' findings. The relevant incentives are not isolated to institutions for scholarly communication, such as editorial policies, for-profit presses, and peer review processes. These institutions are entangled with others, such as research funding, equipment manufacturing

and distribution, organizational policies, hiring and promotion practices, and research assessment exercises. Scientists have a choice about what to publish but not *whether* to publish, and that's because the material resources needed to continue working as scientists are on the line. Thus, explanations based on the economy of recognition are incidental to explaining why it's a competitive strategy to maximize publications.

But, of course, scientists want to produce *true* findings (or accurate, or explanatorily adequate). Feminist philosophers of science, especially those concerned with the critique of science, have long been attentive to the ways in which decisions about what to pursue, and how, can matter to our epistemic appraisals of scientific results (Wylie 1996; Lloyd 2006). More recent models of decision making in inductive risk or practical underdetermination scenarios highlight value trade-offs with respect to relevant epistemic and non-epistemic risks, in the context of an individual scientist's decisions (Douglas 2000; Anderson 2004). This wedds well with a focus on the consequences of incentives: scientists must make choices *viz-a-viz* productivity norms. If they conform to productivity norms to avoid the risks of defying them, then we can explain their choices as a risk mitigation strategy, where values are traded off in light of assessing the relevant risks.

Yet, such an explanation is also only incidental. Again: there is no choice for scientists with respect to *whether* to produce papers. Interactions between the recognitional and material economies of science shape when, where, and how the value trade-offs depicted in inductive risk scenarios arise. This is another sense in which the material and social organization of science constitutively produces epistemically sub-optimal outcomes. MTV has the advantage of recognizing that a distinct *kind* of values, background values, is implicated here. The dominant view in the literature about values in science seems to be that values that can stand in causal or justificatory relations with epistemic judgments must be some kind of mental phenomenon in the mind of a scientist, like presuppositions, or preferences, or attitudes. By contrast, background values are constitutive of epistemic judgments in the sense that they structure which problem spaces are (materially) viable for exploration, when they are explored, and to what extent, since the practices by which these

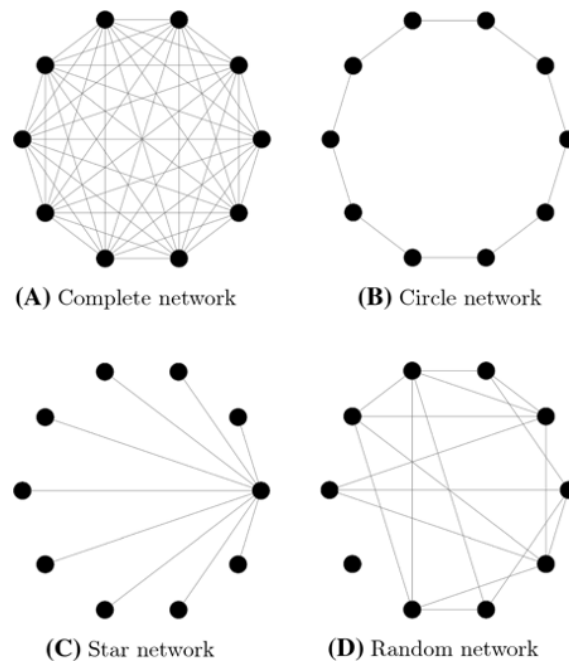
activities proceed are institutionalized practices. It is a distinct advantage of the MTV that it recognizes this broader field of value commitments at stake in science.

The individualistic and cognitivist models considered above are supplementary to explaining the epistemic implications of how scientific communities are organized socially. The MTV offers a more concrete and direct account of the relationship between publish-or-perish and epistemic failures. In addition to addressing open questions about the replication crisis and providing a needed revision to the Mertonian sociological picture, it also has important social epistemological implications, specifically for recent attempts to model scientific communities as networks of Bayesian agents (Zollman 2010, Šešelja 2021). Social epistemologists have been transparent about agent-based network models being highly idealized. It remains a very open question how to de-idealize them (following Knuuttila and Morgan 2019) in a way that would connect concretely with what we already know about the sociological features of scientific communities and practices (cf. Gerson 2013). Individual scientists participate in the institutions that scaffold scientific practices (e.g., peer-review of publications), and those institutions structure how scientists group together (e.g., into specialties). But institutions are not reducible either to groupings of agents or to other organizing structures like communication networks. This suggests that a minimal formal model of a scientific community is at least a three-body problem: both networks and institutions structure groupings of agents.

Implications for Theorizing about Epistemic Agents

The philosophical task of the MTV is to take seriously the sociological and material conditions constituting scientists' epistemic agency. There is risk in failing to take up this task: a risk of committing ourselves to continue theorizing about inhuman entities and pursuing non-empirical studies of science. Our usual theoretical resources for modeling scientific decision-making consist of cognitive models depicting individual agents as bloodless bundles of motivations, beliefs, knowledge, preferences, and other mental states (though, usually not simultaneously). This is especially the case in social epistemology, and especially in projects modeling scientists as simple Bayesian algorithms (i.e., Bayes' theorem plus a fixed updating criterion) that exchange information with other nearby algorithms. What kind of 'agents' are these?

They are not human and barely machine; they are nameless, faceless, and nearly limitless in time and space; and their relations with one another are through anonymous, unmediated channels of exchange of a single type of information. These are otherworldly entities, not even brains in vats. They might as well be Bayesian angels. And the risk is that inquiries like this will amount to nothing more than the new celestial mechanics.



(Figure 1 from Mohseni and Williams 2021)

Perhaps in some future, an indeterminate amount of time away, these simulations will make contact with the corporeal circumstances of embodied inquirers. What might this look like, and how long are we willing to wait for it? Imagine a group of people standing in a circle facing outward. Much available sensory information has been blocked. They wear blindfolds, noise-cancelling headphones, mittens, and thick clothing so that they cannot see, hear, feel, or otherwise be aware of there being anyone else around them. They are asked to train their minds on visualizing the value of a single probabilistic credence. After a period of time, a member is chosen at random to say their credence aloud and the utterance is piped into the headphones of select participants, after which everyone is asked to calculate a new credence to be displayed on

their mental billboard according to their pre-assigned updating criterion. In what sense are these individuals *agents*, let alone scientists? In what sense are they *connected* to each other in a *network*? In what sense is this *communication*? This might seem a rather brute way to depict a potential de-idealization. But one has to start somewhere, since no one has yet tried (as of my writing this), despite claims about the policy relevance of simulation results (cf. O'Connor 2019).

Feminist philosophies in the 20th and 21st centuries have consistently pushed for the social and material re-embodiment of reasoning agents, such as in feminist standpoint theories (cf. Harding 1991). The MTV follows this tradition by assuming as a starting point for theorizing that any epistemic agent is constituted, in part, by the institutions in which they are embedded, such that they have commitments (of various kinds) to the values constitutive of institutionalized practices. These agents are not only flesh and blood; they have histories, rich social relations with one another, and something important at stake in those social relations (i.e., their livelihoods). MTV points us toward the social features that organize collective life in science and the material conditions that constrain scientific work. It aims to show how we might understand scientific epistemology from that perspective: having an integrated understanding of the social, political, and material economy of scientific communities, from which epistemological questions and theories flow. The hope is that this more sociologically robust depiction of scientific communities opens up new territory for exploring the relationship between scientific practices and institutions, for understanding how power operates in and through scientific work, and for conversations between philosophy of science, science policy, and political economy.

References

- Alexandrova, Anna (2018) "Can the Science of Well-Being Be Objective?" *The British Journal for the Philosophy of Science* 69, 421-445.
- Anderson, Elizabeth (2004) "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce", *Hypatia* 19, 1-24.
- Douglas, Heather (2000) "Inductive Risk and Values in Science", *Philosophy of Science* 67, 559-579.

Elliott, Kevin C.; Rebecca Korf (2024) "Values in Science: What Are Values, Anyway?" *European Journal for Philosophy of Science* 14, 1-24.

Gerson, Elihu M. "Institutions and Repertoires" [unpublished manuscript]

Gerson, Elihu M. (2013) "Integration of Specialties: An Institutional and Organizational View" *Studies in History & Philosophy of Science Part C* 44, 515-524.

Guala, Francesco (2016) *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton: Princeton University Press.

Haraway, Donna (1988) "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective", *Feminist Studies* 14, 575-599.

Harding, Sandra (1991) *Who's Science? Who's Knowledge? Thinking from Womens' Lives*. Ithaca: Cornell University Press.

Haslanger, Sally (2022) "Failures of Methodological Individualism: The Materiality of Social Systems", *Journal of Social Philosophy* 53, 512-534.

Heesen, Remco (2018) "Why the Reward Structure of Science Makes Reproducibility Problems Inevitable", *Journal of Philosophy* 115, 661-674.

Hughes, Everett C. (1942) "The Study of Institutions", *Social Forces* 20, 307-310.

Kitcher, Philip (1990) "The Division of Cognitive Labor", *Journal of Philosophy* 87, 5-22.

Knuuttila, Tarja; Mary S. Morgan (2019) "Deidealization: No Easy Reversals", *Philosophy of Science* 86, 641-661.

Korbmacher, Max et al. (2023) "The Replication Crisis Has Led to Positive Structural, Procedural, and Community Change" *Communications Psychology* 1, 1-13.

Lloyd, Elizabeth (2006) *The Case of the Female Orgasm: Bias in the Science of Evolution*. Cambridge: Harvard University Press.

Merton, Robert K. (1957) "Priorities in Scientific Discovery: A Chapter in the Sociology of Science", *American Sociological Review* 22, 635-659.

Mohseni, Aydin; Cole R. Williams (2021) "Truth and Conformity on Networks", *Erkenntnis* 86, 1509-1530.

Nosek, Brian et al. (2022) "Replicability, Robustness, and Reproducibility in Psychological Science", *Annual Review of Psychology* 73, 719-748.

O'Connor, Cailin (2019) *The Origins of Unfairness: Social Categories and Cultural Evolution*. Oxford: Oxford University Press.

- Open Science Collaboration (2015) “Estimating the Reproducibility of Psychological Science,” *Science* 349, 4716.
- Romero, Felipe (2017) “Novelty vs. Replicability: Virtues and Vices in the Reward System of Science”, *Philosophy of Science* 84, 1031-1043.
- Rouse, Joseph (2002) *How Scientific Practices Matter: Reclaiming Philosophical Naturalism*. Chicago: University of Chicago Press.
- Rudner, Richard (1953) “The Scientist Qua Scientist Makes Value Judgments” *Philosophy of Science* 20, 1–6.
- Šešelja, Dunja (2021) “Exploring Scientific Inquiry via Agent-Based Modeling” *Perspectives on Science* 29, 537-557.
- Strevens, Michael (2011) “The Role of the Priority Rule in Science”, *Journal of Philosophy* 100, 55-79.
- Turner, Stephen (1994) *The Social Theory of Practices: Tradition, Tacit Knowledge, and Presuppositions*. Cambridge: Polity Press.
- Ward, Zina (2021) “On Value-Laden Science” *Studies in History and Philosophy of Science Part A* 85, 54-62.
- Wylie, Alison (1996) “The Constitution of Archaeological Evidence: Gender Politics and Science” in: Peter Galison and David J. Stump (eds.) *The Disunity of Science*, Stanford: Stanford University Press, 311-343.
- Zollman, Kevin J. S. (2010) “The Epistemic Benefits of Transient Diversity” *Erkenntnis* 72, 17-35.