Hurried Science and the Tracking Problem

Finnur Dellsén

Forthcoming in a special issue of *Inquiry*, "Science and democracy in times of crises"

(This is a preprint; please cite the published version once available.)

Abstract: In times of crisis, science tends to move fast. However, speeding up science often involves deliberate compromises to the scientific norms that are meant to guarantee the reliability of scientific results in normal circumstances. Call this *hurried science*. Some of the compromises involved in hurried science can be kept track of and later revisited—say, by conducting a new study with larger sample sizes and proper randomization. However, I argue that hurried science may also involve *untracked* compromises to reliability-enhancing scientific norms, in which the negative epistemic effects of speeding up scientific research are not properly noticed or understood. I further argue that these are considerably more worrying for the overall reliability of scientific research, insofar as they are liable to lead both scientists and the general public to overestimate the trustworthiness of scientific results.

1. Introduction

In times of crisis, scientific research is often pursued with increased urgency. Novel results may be needed at short notice so as to guide decisions concerning how best to respond to, and intervene in, the crisis at hand. Under such circumstances, scientists are routinely encouraged and incentivized to speed up their ordinary research process. However, accelerating scientific research will often involve 'cutting corners', i.e. deliberate violations of the norms that are meant to ensure the reliability of scientific research under normal conditions. This is, roughly, what I'll call *hurried science*.

There are obvious benefits to hurried science, especially in times of crisis. For the purposes of making decisions and interventions, what matters most is often to have relevant information at the right moment—even if that information is somewhat less precise or reliable than it could be. So, we definitely need hurried science. With that said, it should also be clear that hurried science can be problematic, insofar as it involves violations of the reliability-enhancing norms of routine science. But what exactly are the potential problems with hurried science? Are these problems equally serious? If not, which problems of hurried science should we take most seriously? And are the problems with hurried science unavoidable, or could a more careful approach to speeding up scientific research circumvent at least some of them?

This paper aims to make headway on these questions—and, more broadly, to contribute to a small but burgeoning conversation in philosophy of science about the problems and pitfalls of speeding up scientific research (see, e.g., Stengers 2018; Stegenga 2024).[1] I'll argue that some instances of hurried science are relatively unproblematic, since the violations of reliability-enhancing norms involved are easily *tracked*. In other instances, however, hurried science is much more problematic, viz. when the violations of reliability-enhancing norms are *untracked*. Roughly speaking, a norm-violation involved in hurried science counts as tracked just in case relevant scientists are aware that the norm was violated and understand how this negatively affects the research in question. Untracked norm-violations are particularly problematic because they are liable to mislead researchers into thinking that a scientific result is more reliable than it actually is.

I will proceed as follows. In section 2, I explain in more detail what I mean by 'hurried science', and how this differs from two similar concepts in the recent literature, viz. 'urgent science' and 'fast science'. In section 3, I introduce what I call the Tracking Problem, which roughly arises whenever hurrying science leads to untracked violations of reliability-enhancing scientific norms. Sections 4 and 5 build on this by examining two general ways in which the Tracking Problem may arise in scientific practice. In particular, these sections discuss two reliability-enhancing norms of routine science—what I'll call the Critique Norm and the Consideration Norm—and examines in detail cases in which these norms were violated in untracked ways. Section 6 concludes by considering when and how it's possible to speed up scientific research without falling foul of the Tracking Problem.

2. Urgent, Fast, and Hurried Science

My aim in this section is to explain in more detail what I mean by 'hurried science', which (for now) may roughly be understood as scientific research that's been sped up in ways that require violations of methodological norms. I'll start by comparing 'hurried science' with two similar notions in the recent literature: Shaw's 'urgent science' and Stegenga's 'fast science'. Although these other notions are clearly closely related to 'hurried science', each was introduced to capture a slightly different range of phenomena than what I will discuss in the following sections.

Shaw's notion of 'urgent science' applies to research proposals. His concern is with how such research proposals should be evaluated in terms of the extent to which they're worth pursuing, i.e. their pursuitworthiness. His point is, roughly speaking, that

[1] In addition, scientists themselves have expressed various concerns about what I call hurried science, especially in the context of the COVID-19 pandemic (see, e.g., Glasziou et al. 2020; Ioannidis 2020; London and Kimmelman 2020; Leite and Diele-Viegas 2021).

some research proposals might be pursuitworthy primarily because they promise to deliver results that can be relied upon relatively soon, or at least within some specified period of time. Shaw's definition of 'urgent science' is meant to capture this: "A research proposal is urgent iff there is a practical or moral reason to need a result within a specified timeline and the research can realistically be carried out within that timeframe" (Shaw 2022a, 108, see also 2022b). Shaw helpfully contrasts these types of research proposals with those whose pursuitworthiness does not depend on the research being completed within any particular period of time. Shaw refers to this as 'luxury science', and defines it as follows: "A research proposal is luxurious iff it has no expected timeline for returning particular results" (Shaw 2022a, 108).

A key difference between *urgent* and *hurried* science is that the former applies to research *proposals* whereas the latter—as we shall see in more detail below—applies to research *processes*. To see clearly the difference here, note first that a research proposal may be urgent even though the process through which research on it is done is not hurried in any way. The research itself may be done in the normal, slow and unhurried manner, even though the topic of that research is very urgent indeed. Moreover, a hurried research process may not address an urgent research proposal. Rather, that research proposal may be luxurious—but, for whatever reason, the research process has nevertheless been sped up or hurried. For instance, an archaeological excavation may be sped up or hurried even though the relevant research proposal has no timeline for returning results—the researchers may simply be speeding up or hurrying their research because they are impatient, especially ambitious, or particularly eager to work on other projects. In sum, then, the urgency of a research proposal is neither a necessary nor a sufficient condition for fast or hurried science.

With that said, it seems highly plausible that urgent research proposals will disproportionally be researched in a hurried manner. If a research proposal is sufficiently urgent, that may well lead researchers to speed up their research even if doing so involves violating methodological norms. So, we should expect to find a strong *correlation* between urgent and hurried science. However, since my discussion of the problematic aspects of hurried science apply even in cases which don't strengthen that correlation, I will not be specifically concerned with urgent science in what follows.

Let us move on, then, to Stegenga's notion of 'fast science'. This notion is meant to apply to research processes, rather than research proposals. Contrary to what one might think, however, 'fast science' is not simply meant to refer to scientific research that occurs at a high, or higher-than-usual, pace.[2] Rather, says Stegenga, "what distinguishes fast science from routine science is the violation of reliability-enhancing principles and practices of routine science (whatever those are), in response to a supreme emergency"

---

[2] This, by contrast, seems to be at least part of what Stengers (2018) is cautioning against in her extended argument for 'slow science'.

(Stegenga 2024, 10).[3] As this makes clear, fast science is distinguished from routine science by its involving (i) violations of what I shall simply call *reliability-enhancing norms*, and (ii) responding to a supreme emergency, such as the spread of COVID-19.[4]

Fast science, thus defined, is quite close to what I'll define as hurried science. In particular, hurried science also involves violations of reliability-enhancing norms. Indeed, this will be crucial in what follows. However, my notion of 'hurried science' is broader notion insofar as it does not require that these violations have come about as a response to a supreme emergency. Consider, for instance, that reliability-enhancing norms may be violated because the scientists in question are hoping to be credited with being the first to have made some discovery. A case in point may be when two electrochemists, Martin Fleishman and Stanley Pons, announced in 1989 that they had achieved cold fusion on the basis of a small tabletop experiment which had not been subjected to peer review and later turned out to be involve methodological flaws and experimental errors (Huizenga 1994). Fleishman and Pons clearly sped up their research and announcement thereof in order to secure priority of their 'discovery', but they did not do so in response to a supreme emergency.

My focus on this broader notion is due mainly to the fact that the problems for hurrying scientific research that I focus on below are not restricted to violations of reliability-enhancing norms that occur in response to supreme emergencies. Stegenga's paper, by contrast, is concerned with issues that only arise when there is a supreme emergency to which scientists are responding. Insofar as my 'hurried science' is simply a broader notion than Stegenga's 'fast science', much or all of what I argue below is true of hurried science will also be true of fast science. In particular, if hurried science faces a Tracking Problem in the way I suggest below, then fast science arguably faces that problem as well. This may make it harder to justify speeding up science even in response to supreme emergencies, insofar as scientific work "ought not depart so radically from

[3] Friedman and Šešelja (2023, 938) define 'fast science' as "application-driven research confronted with an urgent need to accept or reject a certain hypothesis for the purposes of policy guidance, aimed at addressing a significant pending social harm". This is similar to Stegenga's definition of the same term insofar as Friedman and Šešelja's 'significant pending social harm' can be identified with Stegenga's 'supreme emergency', but differs in not requiring any violations of reliability-enhancing norms. What I call 'hurried science', by contrast, *does* require violations of reliability-enhancing norms, and *does not* require addressing any supreme emergency/significant pending social harm. Hence my notion of 'hurried science' is much closer to 'fast science' as defined by Stegenga than as defined by Friedman and Šešelja.

[4] Stegenga sometimes writes as if the speed at which scientific work is carried out is irrelevant to whether something counts as fast science on his definition of the term. In particular, speed is not mentioned in what seems to be the official definition the term (Stegenga 2024, 10). For this reason I do not include speed in this two-item list of what distinguishes fast from routine science. However, I take it that Stegenga's intention may well have been to distinguish a special kind of sped-up science from another, more routine, form of sped-up science. In that case, high speed is still a necessary condition of fast science, but clearly not a sufficient one.

routine science such that the resulting work becomes highly unreliable" (Stegenga 2024, 11).[5]

Having discussed how hurried science differs from urgent science and fast science, let us turn to how it may be captured in a definition. I'll take the following three conditions to be individually necessary and jointly sufficient. First, hurried science is done at a higher-than-usual pace, i.e. faster than scientific research is normally carried out within the relevant research specialization or discipline. Second, it violates some reliability-enhancing norm of routine science. Third, the norm-violations in the previous condition are due to the higher-than-usual pace in the first condition. In brief, then, hurried science may be defined as scientific research that (i) is done at a higher-than-usual pace, (ii) violates reliability-enhancing norms, (iii) where (ii) is due to (i). Even briefer: hurried science violates some reliability-enhancing norms because of its higher-than-usual pace.

What does hurried science look like in practice? In sections 4 and 5, I discuss in considerable detail two actual cases of hurried science, in the course of illustrating the way in which hurried science is often problematic. For now, let me exemplify hurried science with a much simpler, hypothetical case.

Suppose that a particular biomedical lab, Lab A, is under intense pressure to get quick results, e.g. so as to secure funding for the next year. Normally, Lab A would recruit a minimum of 50-200 participants, divide them randomly into a treatment and control group, and follow other standard methodological procedures for randomized controlled trials (RCTs). In this case, however, the head of Lab A decides to speed up the lab's research process by recruiting only 10 participants in total. The study consequently suffers from a very limited statistical power—much lower than the 80-90% power that such studies are generally expected to showcase. This makes any statistical conclusions drawn from it highly dubious, even if—let's say—the lab obtains a result that counts as statistically significant at the 0.05 level.

This is a case of hurried science insofar as the research process (i) was significantly sped up by using 10 as opposed to 50-200 participants, (ii) violated the reliability-enhancing norm within biomedical science of aiming for at least 80-90% power, and (iii) the latter was due to the former. It is not a case of fast science, however, since although this research does violate a reliability-enhancing norm, it doesn't do so in response to a supreme emergency. I also suspect that Shaw would not count the research proposal as urgent science—unless, that is, securing funding to continue running a lab for another year constitutes "a practical or moral reason to need a result within a specified timeline" (Shaw 2022a, 108).

[5] Stegenga refers to this as *the Principle of Similarity*. It is one of Stegenga's two principles for assessing fast science.

3. The Tracking Problem

This section describes an epistemic problem that often (though not always) arises in cases of hurried science. For reasons that will become clear below, I'll refer to this as *the Tracking Problem*. To explain the problem, let me start by considering a case of hurried science in which that problem does *not* arise.

A simple case of that sort can be obtained by expanding slightly on the simple example with which I ended the previous section. So, consider again Lab A's RCT with only 10 participants and limited statistical power. Now let's suppose that the small number of participants, as well as the correspondingly low statistical power, are duly noted when Lab A's results are written up and published. Future research on the treatment in question, as well medical practice, is correspondingly hesitant to place much weight on Lab A's results (even if statistically significant at the 0.05 level)—unless and until the results can be reproduced in a larger and correspondingly more powerful study.

What I wish to emphasize about this case is that the limited statistical power of this RCT, and the adverse effects of these on the trustworthiness of the results, do not go unnoticed. On the contrary, that community is well aware that Lab A's RCT fell short of fully living up to a methodological norm in their field. Furthermore, the community understands well what this means for the reliability of the study's results. Accordingly, there will be no confusion, going forward, about whether this study should be relied upon to the same extent as larger and more powerful RCTs. So, while it's perhaps not ideal that Lab A's study was done in such a hurried manner, the negative repercussions of the relevant norm-violation would seem to be minor at worst.

Although this example is hypothetical, there are many actual cases of this kind. Consider, for instance, the various small studies on COVID-19 treatments that were carried out during the first few months of the pandemic. In an early commentary on this, Glasziou et al. (2020, 1) pointed out that of the studies that were being done on whether hydroxychloroquine treatments at the time, "32 [had] a planned sample size of ≤100, 10 [had] no control group, and 12 [were] comparative but non-randomised." Fortunately, however, scientists were under no illusion regarding the reliability of these studies;[6] and a much larger and more systematic study, with a total of 4716 participants across the treatment and control groups, was subsequently done by the RECOVERY Collaborative Group (2020). This study showed no mortality benefit (after 28 days) to hydroxychloroquine treatment, settling the matter to most scientists' satisfaction.

[6] Some scientists, such as Glasziou et al. (2020) and London and Kimmelman (2020), argued that researchers should instead pool their resources so as to conduct larger and more reliable studies rather than multiple small ones.

In what follows, I will argue that many norm-violations in hurried science are not like this. My argument rests on a distinction between 'tracked' and 'untracked' norm-violations, which in turn is based on the following definition:

> *Tracking (definition)*. Some agents *track* a norm-violation just in case the agents are aware of the norm-violation and understand how it adversely affects the research in question.

This definition relativizes tracking to a group of agents, whose awareness and understanding of the norm-violation is required for the violation to count as tracked. In what follows, however, I will almost exclusively be concerned with whether a given norm-violation is tracked by the scientists who specialize in the type of research in which that violation took place. Accordingly, I'll often simply describe a given norm-violation as *tracked* or *untracked* without explicitly mentioning the agents by whom it is tracked/untracked, but take it as read that those agents are the scientists who specialize in the relevant type of research.[7]

Lab A's norm-violation in its 10-person RCT clearly counts as tracked by this definition, since members of the wider biomedical community who specialize in this type of research are fully aware of Lab A's violation and its effects on the study's reliability. So, we already have an example of tracked norm-violations in hurried science. But what would be an example of an untracked violation of a reliability-enhancing norm? The following two sections contain case studies that serve as detailed examples of such untracked violations. For now, however, let us consider a much simpler hypothetical case that's designed to be as similar as possible to the case of Lab A's RCT except insofar as it involves an untracked rather than tracked violation of a reliability-enhancing norm.

Suppose another biomedical lab, Lab B, is under a similar pressure as Lab A to obtain quick results. The head of Lab B, however, is not tempted to recruit fewer participants than usual; rather, Lab B's researchers soldier on as usual with 50-200 participants per study. However, after randomly assigning participants into treatment group and control groups, numerous patients in the treatment group drop out, leaving the two groups unbalanced in terms of their numbers and various baseline characteristics (e.g., age, gender, education). In order to avoid having to restart the study at this point, the head of Lab B orders their researchers to 're-balance' the two groups by reassigning several members of the control group to the treatment group. Lab B then obtains

[7] It is worth noting that by this definition any norm-violation can count as tracked or untracked, regardless of whether the norm-violation is due to the speeding up of scientific research. For instance, one can easily imagine a scientific norm being violated for financial reasons, as when a lab simply lacks funding to recruit more than 10 participants. Such a norm-violation can be tracked and untracked, of course, depending on whether relevant agents are aware of it and understand its adverse epistemic effects on the research in question.

statistically significant results with adequate power, and promptly publishes its results in a prominent journal—without mentioning the 're-balancing' of participants.

Lab B's research violates a reliability-enhancing norm inherent in the practice of RCTs. By 're-balancing' the participants, the participants no longer count as properly randomized, and the results eventually obtained may well end up being due to an unintentional selection bias. In this case, the norm-violation is *untracked*, because the larger biomedical community will not be aware that this norm was violated—since it was not mentioned in the published article—and will therefore not understand the adverse effects on the epistemic status of Lab B's results. This is clearly a more problematic instance of hurried science than Lab's A's 10-person RCT. After all, Lab B's epistemically dubious results may well form the basis of further research within biomedical science, quite possibly wasting the field's limited research resources on wrongheaded follow-up projects. In addition, Lab B's results may well be relied upon in medical practice, which could involve serious harms to patients.

This suggests that the much more problematic norm-violations in hurried science are those that are untracked.[8] In contrast to tracked violations of reliability-enhancing norms, untracked violations are liable to mislead both researchers and the general public into thinking that a scientific result is more reliable than it actually is. The risk that hurrying scientific research may involve untracked norm-violations is what I shall refer to as 'the Tracking Problem'. More precisely:

> *The Tracking Problem*. The risk that, in a given instance of hurried science, reliability-enhancing scientific norms are violated in untracked ways.

How serious is the Tracking Problem? How commonly is there a risk that the norm-violations involved in a given instance of hurried science are untracked?[9]

In what follows, I will consider two very general reliability-enhancing scientific norms, and show that both of these can and have been violated in untracked ways. This shows that the risk of untracked norm-violations in hurried science is not confined to

[8] Indeed, one might even argue that some or all putative norm-violations are really only genuine norm-violations when they go untracked. On this view, there would be no real norm violated by non-randomized studies, for instance, provided there is sufficient tracking of the fact that the studies were not randomized. (Thanks to [anonymized] for pointing this out.) While I agree with the spirit of this idea—since there might be nothing wrong, all things considered, with violating putative norms in this way—I find it more straightforward to think of scientific norms as not including an escape clause to the effect that a putative norm hasn't been violated if it is tracked. The latter construal of scientific norms would add to their complexity, and make them less similar to the rules and guidelines that working scientists themselves tend to endorse.

[9] I focus in this paper on the risk of untracked violations of reliability-enhancing norms *in hurried science*, i.e. untracked norm-violations that are due to the speeding up of scientific research (see section 2). However, there is a similar risk that norm-violations that come about in other ways, e.g. because of lack of funding, are untracked (see footnote 6). In this sense, there is a more general Tracking Problem for *any* norm-violattions in science—call this the *Generalized Tracking Problem*. I leave discussion of that problem to future work.

some discipline-specific scientific norms, such as the norms around RCTs; rather, the scope of the Tracking Problem is considerably more general. Moreover, in contrast to the norm against 're-balancing' participants in RCTs, the general reliability-enhancing norms I'll consider below are rarely made explicit, e.g. in each discipline's textbooks and discussions of methodological issues. For this reason, scientists may be less alert to violations of these norms, and thus less likely to track them.

4. The Critique Norm

Duhem (1954) famously argued that scientific argumentation often relies heavily on background assumptions with which data are connected to theories and hypotheses. These background assumptions often remain tacit or implicit in the published arguments formulated in published works. Accordingly, these background assumptions are not subject to the same intense scrutiny as are other components in scientific arguments, such as the empirical data and mathematical theorems involved therein. Indeed, recent work in philosophy of science has argued, following Longino's (1990, see also 1994) seminal argument, that such background assumptions often conform to the 'contextual' values that individual scientists have adopted throughout their lives, e.g. their political convictions, ethical commitments, and—problematically—their racial and gender biases (Intemann 2015; Elliott 2017).

However, while the arguments proposed by *individual* scientists will thus often lack a certain kind of objectivity to which science arguably aspires, Longino and others suggest—correctly, in my view—that the same will not generally be true of the arguments endorsed by the scientific community as a whole. In particular, suppose this community contains scientists with a sufficiently wide range of different 'contextual' values, each of whom has the opportunity to criticize the scientific arguments formulated by their fellow scientists. In that case, we can be reasonably confident that any major epistemic issues due to faulty or biased background assumptions will have been weeded out before the arguments in question become endorsed by the scientific community as a whole (Longino 2001; Solomon 2001).

Notice, however, that this manner in which scientific objectivity is obtained at a at the level of scientific communities requires that the individual scientists whose contextual values would conform to background assumptions that are different from one's own have genuine opportunities to formulate and express their criticism once a new argument is introduced. Thus, the process of validating a given scientific claim requires scientists to adhere to:

> *The Critique Norm*. Before accepting[10] a given scientific claim *C*, one must first provide other researchers with the opportunity to scrutinize the operative background assumptions in one's argument(s) for *C*.

This is clearly a reliability-enhancing scientific norm—the violation of which, if due to the speeding up of the scientific process, would thus make for hurried science.[11]

Now, in the abstract it seems plausibly that the Critique Norm would be violated in a great many cases in which science is sped up to any significant extent. If the pace of scientific work done by some particular individual or group is sufficiently high, it is hard to see how other scientists can be given a realistic opportunity to scrutinize the operative background assumptions in that individual's or group's arguments. If so, there is an inherent tension between speeding up scientific research to any significant extent and abiding by the Critique Norm. As we shall see, this can be backed up by at least one detailed case study in which speeding up scientific research seems to have led to a clear violation of the Critique Norm.

Moreover, it also seems plausible in the abstract that violations of this norm will be particularly likely to be untracked. Recall from section 3 that tracking requires not just being aware of the fact that a norm has been violated, but also understanding how its violation adversely affects the research in question. In the case of violating the Critique Norm, it will often not be clear how failing to give other researchers the opportunity to scrutinize one's background assumptions affects the epistemic status of one's own arguments and conclusions. Those background assumptions may well be in good standing; or they may be mistaken or biased in any number of different ways. Unless and until those other researchers are actually provided with the opportunity to scrutinize these background assumptions, it seems that little can be known about what the epistemic effects on one's arguments and conclusions would turn out to be.

Let me substantiate these points with a real case of hurried science in which, I submit, the Critique Norm was clearly violated in an untracked manner. The case concerns a health study conducted in 1978 by the New York State Department of Health on several hundred homes in the neighborhood of Love Canal in the city of Niagara Falls. These homes were thought to be at risk of chemical contamination from an old chemical dumping ground in the middle of the neighborhood, on top of which the local school had been placed. The focus, in particular, was on whether the chemicals were seeping into the ground in Love Canal and causing women to have an abnormal number of

---

[10] I don't mean anything very specific by 'acceptance' here or in what follows. You may think of it simply as the point at which a given scientific claim is widely considered to be secure enough to be acted on, e.g. in a medical intervention or in public policy.

[11] Since I am borrowing the notion of a reliability-enhancing norm from Stegenga's notion of fast science, it is relevant to note that Stegenga (2024, 11) uses "Longino's norm that science should be responsive to criticism" (which I take to be roughly what I'm calling the Critique Norm) as a prime example of a reliability-enhancing norm which can be violated in fast science.

miscarriages. In August 1978, 239 homes were evacuated, and the Department of Health was tasked with investigating whether further evacuations were necessary.

My discussion of this episode relies on a first-hand report by Beverely Paigen (1982). Paigen worked as a cancer researcher at the Department of Health at the time, although she was not directly involved in the department's original investigation into Love Canal homes. Paigen describes the investigation as follows:

> The Health Department initiated health studies of 850 additional homes in the Love Canal neighborhood by distributing questionnaires and taking blood samples for analysis. In early fall of 1978, the department announced the preliminary results of these studies; officials assured the Love Canal residents that the neighborhood was a safe place to live and that the community beyond the homes that had already been evacuated was not at any increased health risk. This announcement was based on data showing that the miscarriage rate in homes beyond the barrier was no higher than elsewhere (Paigen 1982, 29–30).

It should be clear from this description that the Department of Health's investigation exemplified a speeding up of the normal scientific process: the investigation began in August 1978 at the earliest, and concluded in the "early fall" of that same year, with officials offering an assurance that Love Canal was a safe place to live.

But why think that the Department of Health's investigation violated the Critique Norm? Recall that this norm requires giving other researchers opportunities to scrutinize the operative background assumptions in one's scientific arguments before accepting the conclusions of those arguments. In this case, the most salient instance of such 'other researchers' was Paigen herself, whose research at the time focused on differences in susceptibilities to environmental toxins of the sort found in Love Canal. However, presumably due to the haste with which the investigation was conducted, Paigen was not invited to comment on the Department of Health's studies or conclusions before Love Canal residents were assured that their neighborhood was safe. Had she been so invited, the department would almost certainly have reached a different conclusion, as we shall now see.

The Department of Health's arguments were based on two main assumptions. First, the department assumed that any increased miscarriage risk from the toxic waste under Love Canal would mean that the rate of miscarriages in Love Canal would decrease with distance between a peoples' homes and the toxic dumping ground in the center of the neighborhood. This predicted correlation between distance from the dumping ground and miscarriage rates within Love Canal was not observed. Second, the department also assumed that, given such an increased health risk, miscarriages would be more frequent in Love Canal than the 'baseline' miscarriage rate in the United States as measured in a study by Warburton and Fraser (1964). In fact, the observed frequency of miscarriages in Love Canal (16%) were not significantly higher than the 'baseline' miscarriage frequency in the US as measured by Warburton and Fraser (14.7%).

Paigen conducted her own small study of Love Canal, independently of the official Department of Health investigation described above, and without any funding from the department or other sources. Interestingly, Paigen came to quickly reject both assumptions in her own department's case for Love Canal being a safe neighborhood.

With respect to the first assumption, Paigen showed that although rates of miscarriage in Love Canal did not correlate with distance to the dumping ground, they did correlate with location more generally. Specifically, Paigen was able to show how these miscarriage rates were significantly higher in 'wet' locations within Love Canal, where there had previously been stream beds and swales that collected still water from the ground. Since toxic chemicals would plausibly migrate preferentially along these stream beds and swales, rather than indiscriminately migrating through the ground in every direction from the dumping ground, this is precisely what should be expected if the toxic chemicals from the dumping ground were causing miscarriages. By contrast, miscarriage rates should not be expected simply to decrease with the distance from the dumping ground to people's homes, as the Department of Health's official investigation had assumed.

What about the second assumption, viz. that if chemicals from the dumping ground were causing miscarriages, then miscarriages among Love Canal residents should be expected to be significantly more common than the general frequency measured by Warburton and Fraser (1964)? Paigen found several serious issues with this assumption, most of which had to do with whether the women who participated in Warburton and Fraser's study could really be assumed to be sufficiently representative of women in Love Canal to serve as a 'control' for the Department of Health's investigation. In particular, the women in Warburton and Fraser's study had all previously either given birth to twins or a child with birth defects. Moreover, these women were all economically disadvantaged and had little or no access to prenatal care. Indeed, Paigen noted that the miscarriage rate of Love Canal women *before* they moved to the neighborhood was much lower than this 14.7% 'baseline', viz. 8.5%. For these reasons, Paigen argued that it was unreasonable to assume that the 14.7% miscarriage rate measured by Warburton and Fraser represented any sort of baseline with which the 16% Love Canal miscarriage rate could be meaningfully compared.

Taking stock, Paigen criticized the Department of Health's investigation of Love Canal for making two highly questionable assumptions (which, as it turns out, were both false). It seems clear that these assumptions would not have been relied upon if Paigen had had the opportunity to scrutinize them before the Department of Health concluded its investigation of Love Canal, announcing to its residents that the neighborhood was a safe place to live. It therefore seems that the Department of Health's failure to live up to the Critique Norm undermined, at least to some extent, the reliability of its conclusions. Since this was not acknowledged by the Department of Health in any way, let alone understood by relevant scientists to adversely affect the reliability of the department's investigation, this violation was moreover untracked.

Indeed, it is arguably the lack of tracking that makes the Department of Health's conduct so problematic. The team of researchers who conducted the official investigation were presumably under immense pressure to deliver results in a timely manner. As such, these researchers may reasonably have considered it justifiable to flout the Critique Norm on this particular occasion, announcing the results of their investigations before other researchers had had an opportunity to scrutinize their background assumptions. Had they accompanied such an announcement with an acknowledgement of the epistemic limitations of their investigation, regarding in particular the lack of scrutiny from other researchers, it would have been hard to find any fault with their conduct. What is problematic in this case, then, is not the hurrying of scientific research itself, or indeed the norm-violations involved therein, but the fact that these norm-violations were not tracked.

5. The Consideration Norm

Much of scientific reasoning is comparative, consisting of an explicit or implicit comparison between alternative theories of some phenomena. A theory T_1 is then accepted because T_1 compares favorably to a range of alternative theories T_2, T_3, etc., that have also been proposed at some point. This comparative nature of scientific reasoning is particularly salient in *Inference to the Best Explanation* (IBE), according to which an explanatory theory may be accepted only if it provides an overall better explanation of relevant evidence than other available rival theory (Harman 1965; Lipton 2004; Dellsén 2024). However, many other accounts of scientific reasoning—including variants of the Bayesian account—similarly take scientific reasoning to be comparative in this way (at least much of the time). For instance, Salmon (1990) famously argues that scientific reasoning often only involves probabilistic comparisons between available theories rather than estimations of the absolute values of these theories' probabilities (see also, e.g., Royall 2000; Sober 2008).

A well-known practical issue for comparative scientific reasoning of this sort concerns the fact that scientists are ever only able to formulate, let alone systematically compare, a highly limited set of all the possible alternative theories of the phenomena (Sklar 1981; Stanford 2006; Wray 2021). Now, suppose that the set of theories that scientists have actually formulated and compared contains only mistaken theories, or is biased in some way. If so, then the conclusion of a comparative piece of scientific reasoning will clearly be similarly mistaken, biased, or otherwise epistemically problematic. In the context of IBE, van Fraassen (Fraassen 1989) famously refers to this as 'the problem of the bad lot'—but it's not hard to see that the problem generalizes to any comparative evaluation of scientific theories (Okruhlik 1994; Dellsén 2020). In what follows, I will follow Lipton (1993) in referring to this as *the underconsideration problem* (see also Khalifa 2010; Dellsén 2017).

In practice, this risk of comparing theories within a 'bad lot' may often be significantly mitigated—if perhaps never eliminated entirely. Dellsén (2021) discusses how scientists are in fact hesitant to accept a scientific theory on the basis of a comparative explanatory evaluation, such as the one involved in IBE, unless the theory has gone through a process he calls *explanatory consolidation*. A critical part of this process consists in scientists making failed attempts to formulate better alternatives, which gradually makes it more plausible that the best explanatory theory currently on offer is better than any other that could be formulated. Crucially, this generally does not happen overnight; rather, it involves a significant passing of time during which scientists work to formulate alternative explanatory theories of the phenomenon in question.

The fact that scientists refuse to accept, on comparative grounds, theories that have not gone through a process like explanatory consolidation suggests that they follow a (presumably implicit) norm to that effect:

> *The Consideration Norm*. Before accepting a given scientific claim *C* on comparative grounds, one must first formulate and actively consider a range of alternative theories that may better account for the data than *C*, and/or provide others with the opportunity to do so.

Like the Critique Norm, the Consideration Norm is also clearly reliability-enhancing: following it increases the reliability of scientific results (in this case, the results obtained via comparative reasoning such as IBE). The violation of this norm of routine science, if due to the speeding up of the scientific process, would thus also make for hurried science.

It seems clear that the Consideration Norm will be violated in many, though presumably not all, cases in which science is sped up to any significant extent. After all, conforming to this norm will often involve imposing a non-trivial temporal gap between the initial piece of comparative reasoning, in which T_1 is compared to T_2, T_3, etc., and the eventual acceptance of T_1. It is during this time that attempts to formulate alternative theories are made, either by the advocates of T_1, or indeed by its opponents (who may be better placed to spot its weaknesses). When science is sped up, this temporal gap gets squeezed so that the Consideration Norm becomes more difficult to follow. There is thus an inherent tension between speeding up scientific research to any significant extent and conforming to the Consideration Norm.

Can violations of the Consideration Norm be tracked? In principle, the answer is surely yes. It is not impossible for scientists to acknowledge and understand how a particular scientific theory is less reliable due to a lack of time during which alternative theories could be formulated and actively considered. In practice, however, it seems that working scientists are generally not particularly well placed to understand the adverse epistemic effects of the underconsideration problem that the Consideration Norm is designed to counter. Stanford (2006) discusses several historical cases in which the most prominent theorists of the mechanism of biological heredity of the 19th century—Darwin, Galton, and Weisman—all confidently espoused radically false theories on the grounds that no more plausible theories could be developed. These cases suggest that scientists themselves are not particularly adept at understanding when the underconsideration of alternative theories adversely affects the epistemic status of the theories they themselves favor.

To substantiate and illustrate these points, let us turn to another example from the recent history of science.[12] On May 18th, 1980, the active volcano Mount St. Helens in Washington state erupted in a particularly catastrophic manner. The north flank of the mountain suddenly collapsed in a huge landslide, prompting the volcano to erupt in a massive lateral blast towards the mountain's north. The pyroclastic flow from the volcano—consisting among other things of volcanic gases, ash, and rock—travelled laterally at over 1000 km/h, blowing down 600 km^2 of forest in around 30 seconds. Thousands of animals and 57 people were killed, many of whom were outside of the official evacuation zone that had been set up to protect the nearby residents and forestry workers.

Geologists working for the United States Geological Survey (USGS) had been tasked with predicting the most likely eruption scenarios when Mount St. Helens first started showing signs of volcanic activity approximately two months earlier. These geologists relied heavily on a recent study of Mount St. Helens by Crandell and Mullineaux (1978), in which a potential north-flank landslide, causing the unusual lateral eruption (as opposed to the more common vertical eruption) was not among the scenarios considered. With that said, the possibility of some sort of north-flank landslide was discussed by USGS scientists on site, following clear signs of swelling on the mountain's north side (Lipman et al. 1981). But this scenario appears not to have been considered systematically enough to provide a detailed prediction about its size, or its effects on the subsequent eruption (Miller et al. 1981; Tilling et al. 1990). Accordingly, the size and location of the evacuation zone around Mount St. Helens was not adequate to prevent significant loss of human as well as non-human lives when the volcano finally erupted.

[12] In what follows, I rely primarily on Lipman and Mullineaux (1981) and Tillman et al. (1990); but see also Olson (2016).

USGS geologists were clearly under immense time pressure to deliver reliable predictions regarding how, if at all, Mount St. Helens would erupt. Due to these time constraints, they were not able to actively consider a significant range of possible eruption scenarios (nor were they able to provide other scientists with the opportunity to do so). So the speeding up of USGS's investigation into Mount St. Helens caused them to violate the Consideration Norm, making this an instance of hurried science. Moreover, since the USGS did not publicly acknowledge that they had violated this norm, or indeed themselves fully understand what effects this might have on the results of their investigations, the norm-violation in question counts as untracked.

Again it seems clear that the lack of tracking is what makes this case into a particularly problematic instance of hurried science. It is hard to see how USGS geologists could have conformed to the Consideration Norm in the spring of 1980, given how briefly they were aware of volcanic activity in Mount St. Helens at all, and the even briefer period during which they were aware of the bulge in the volcano's north flank. So, in this case, it seems to have been more or less unavoidable that some hurrying of scientific research would take place. By contrast, the norm-violations involved in this instance of hurried science could have been much more diligently tracked, in that USGS scientists could have warned the local residents and forestry workers that not all eruption scenarios had been actively considered, and explained how this undermines the epistemic reliability of their predictions . So, as in the previous case of the NY State Department of Health's investigation into Love Canal, what arguably proves problematic here is not simply the speeding up of scientific research itself, or even the norm-violation caused by it, but rather that these norm-violations went untracked.

6. Conclusion: Tracking while Hurrying

Hurried science is when scientific research is sped up in such a way that reliability-enhancing norms get violated in the process. While hurrying science is not per se particularly problematic, it becomes much more so when relevant scientists and/or members of the general public are not aware of these norm-violations, or fail to understand how they negatively influence the reliability of the research in question. Put differently, what's most problematic about many instances of hurried science is not the norm-violations as such, but the fact that these norm-violations were not tracked.

This risk of untracked norm-violations in hurried science is exemplified by the two case studies discussed above, concerning the NY State Department of Health's investigation into toxic chemicals in Love Canal and the USGS's research into Mount St. Helens's volcanic eruption scenarios. These case studies suggest that the risk of untracked norm-violations in hurried science—i.e., what I've called the Tracking Problem—is quite general. After all, the reliability-enhancing norms violated in these cases—i.e., what I've called the Critique Norm and the Consolidation Norm—are of a very general nature. Moreover, since these norms are rarely made explicit amongst scientists themselves, they are considerably less likely to be tracked than violations of explicit norms, such as those concerning how to carry out an RCT.

As with most risks, however, that which is involved in the Tracking Problem can be circumvented and mitigated through appropriate measures. The most obvious way to circumvent the Tracking Problem is to avoid hurrying scientific research in the first place. If science isn't sped up in ways that involve violations of reliability-enhancing norms, there is no hurrying of scientific research at all. *A fortiori*, there would be no risk of untracked norm-violations in hurried science. Note that this is not necessarily to require that scientific research is not sped up at all; rather, it's to require that science is sped up only in ways that wouldn't involve violating any reliability-enhancing norms. For instance, one might imagine that the NY State Department of Health's investigation into toxic chemicals in Love Canal could have been sped up to some extent while still giving Beverly Paigen—as well as other scientists outside of the department's official investigation team—opportunities to scrutinize the background assumptions with which the team operated.

Presumably, however, there are limits to how much scientific research can realistically be sped up without thereby violating some reliability-enhancing norms. Indeed, some such norms—such as the Consideration Norm—may practically require scientific research to move at a slower pace, insofar as the process of formulating and actively considering alternative theories simply is not the sort of process that can easily be accelerated. At a practical level, then, speeding up scientific research may almost inevitably involve some violations of reliability-enhancing norms. If (and to the extent that) this is so, it may not be practically feasible to circumvent the Tracking Problem by requiring that scientific research is sped up only in ways that don't involve any norm-violations.

This suggests that a more promising approach to the Tracking Problem would be to ensure that, when reliability-enhancing norms are (perhaps inevitably) violated in sped-up science, these violations are properly tracked. To put it crudely, the Tracking Problem can be solved with more diligent tracking. This would require working scientists to know which reliability-enhancing norms are operative in routine scientific research, to be alert to instances in which those norms are violated, and to understand how such violations adversely affect the research in question. In the case of the USGS's research on Mount St. Helens in the spring of 1980, for instance, this would imply a greater sensitivity on the part of USGS geologists to the epistemic effects of their failure to actively consider the lateral eruption scenario. At a practical level, this might involve USGS geologists communicating to local residents that only a limited number of eruption scenarios had been considered, and that it may therefore be advisable to show extra caution until a greater range of scenarios could be explored.[13]

In short, if hurrying scientific research is unavoidable—as it often is—then let us at least make sure that the norm-violations which undermine the research's reliability are properly tracked. Philosophers of science have an important role to play here, since the reliability-enhancing norms whose violations need to be tracked are often implicit in scientific practice rather than explicitly stated in official scientific communications. This is certainly the case for the Critique Norm and the Consolidation Norm, both of which are drawn from discussions amongst philosophers of science aiming to describe and understand scientific practice. It's plausible that there are other similarly implicit reliability-enhancing scientific norms that are regularly violated as a result of speeding up scientific research. Philosophers of science are particularly well placed to identify these norms, describe how they can be (and have in fact been) violated in untracked ways, and finally explain how such untracked norm-violations may be avoided in the future.[14]

References

Crandell, Dwight Raymond, and Donal Ray Mullinieaux. 1978. *Potential Hazards from Future Eruptions of Mount St. Helens Volcano, Washington*. U.S. Geological Survey. https://doi.org/10.3133/b1383C.

Dellsén, Finnur. 2017. "Reactionary Responses to the Bad Lot Objection." *Studies in History and Philosophy of Science Part A* 61: 32–40. https://doi.org/10.1016/j.shpsa.2017.01.005.

Dellsén, Finnur. 2020. "The Epistemic Impact of Theorizing: Generation Bias Implies Evaluation Bias." *Philosophical Studies* 177 (12): 3661–78.

Dellsén, Finnur. 2021. "Explanatory Consolidation: From 'Best' to 'Good Enough.'" *Philosophy and Phenomenological Research* 103 (1): 157–77. https://doi.org/10.1111/phpr.12706.

Dellsén, Finnur. 2024. *Abductive Reasoning in Science*. Cambridge University Press.

Duhem, Pierre Maurice Marie. 1954. *The Aim and Structure of Physical Theory*. no. 1. Princeton University Press.

[13] If USGS geologists had communicated this to local residents, they would effectively have ensured that their violation of the Consideration Norm was tracked not only by themselves and other scientists, but also at least to some extent by the local residents themselves.

[14] I am grateful for helpful feedback on drafts of this paper from Tina Firing, Ísak Andri Ólafsson, Jacob Stegenga, Oscar Westerblad, and an anonymous referee. I am also grateful for excellent comments and question from the audience at the "SciDem in Iceland" workshop, organized by the University of Oslo.

Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190260804.001.0001.

Fraassen, Bas C. Van. 1989. *Laws and Symmetry*. Clarendon Press.

Friedman, Daniel C., and Dunja Šešelja. 2023. "Scientific Disagreements, Fast Science and Higher-Order Evidence." *Philosophy of Science* 90 (4): 937–57.

Glasziou, Paul P., Sharon Sanders, and Tammy Hoffmann. 2020. "Waste in Covid-19 Research." *BMJ*, May 12, m1847. https://doi.org/10.1136/bmj.m1847.

Harman, Gilbert H. 1965. "The Inference to the Best Explanation." *Philosophical Review* 74 (1): 88–95.

Huizenga, John R. 1994. *Cold Fusion: The Scientific Fiasco of the Century*. Repr. Oxford University Press.

Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–32.

Ioannidis, John P. A. 2020. "Coronavirus Disease 2019: The Harms of Exaggerated Information and Non-evidence-based Measures." *European Journal of Clinical Investigation* 50 (4): e13222. https://doi.org/10.1111/eci.13222.

Khalifa, Kareem. 2010. "Default Privilege and Bad Lots: Underconsideration and Explanatory Inference." *International Studies in the Philosophy of Science* 24 (1): 91–105. https://doi.org/10.1080/02698590903467135.

Leite, Luciana, and Luisa Maria Diele-Viegas. 2021. "Juggling Slow and Fast Science." *Nature Human Behaviour* 5 (4): 409–409. https://doi.org/10.1038/s41562-021-01080-1.

Lipman, P. W., J. G. Moore, and D. A. Swanson. 1981. "Bulging of the North Flank before the May 18 Eruption-Geodetic Data." In *The 1980 Eruptions of Mount St. Helens, Washington*, edited by Peter W. Lipman and Mullineaux Donal R., vol. 1250. U.S. Geological Survey Professional Paper. U.S. Geological Survey. https://doi.org/10.3133/pp1250.

Lipman, Peter W., and Donal R. Mullineaux, eds. 1981. *The 1980 Eruptions of Mount St. Helens, Washington*. Vol. 1250. U.S. Geological Survey Professional Paper. U.S. Geological Survey. https://doi.org/10.3133/pp1250.

Lipton, Peter. 1993. "VI*—Is the Best Good Enough?" *Proceedings of the Aristotelian Society* 93 (1): 89–104.

Lipton, Peter. 2004. *Inference to the Best Explanation*. Routledge.

London, Alex John, and Jonathan Kimmelman. 2020. "Against Pandemic Research Exceptionalism." *Science* 368 (6490): 476–77. https://doi.org/10.1126/science.abc1731.

Longino, Helen. 2001. *The Fate of Knowledge*. Princeton University Press.

Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Vol. 25. no. 2. Princeton University Press.

Longino, Helen E. 1994. "Gender and Racial Biases in Scientific Research." In *Ethics of Scientific Research*. Rowman & Littlefield.

Miller, C. D., D. R. Mullinieaux, and D. R. Crandell. 1981. "Hazards Assessments at Mount St. Helens." In *The 1980 Eruptions of Mount St. Helens, Washington*, edited by Peter W. Lipman and Donal R. Mullineaux, vol. 1250. U.S. Geological Survey Professional Paper. U.S. Geological Survey. https://doi.org/10.3133/pp1250.

Okruhlik, Kathleen. 1994. "Gender and the Biological Sciences." *Canadian Journal of Philosophy, Supplementary Volume* 20 (sup1): 21–42.

Olson, Steve. 2016. *Eruption: The Untold Story of Mount St. Helens*. Norton & Company.

Paigen, Beverly. 1982. "Controversy at Love Canal." *The Hastings Center Report* 12 (3): 29. https://doi.org/10.2307/3561826.

Royall, Richard M. 2000. *Statistical Evidence: A Likelihood Paradigm*. Repr. Monographs on Statistics and Applied Probability 71. Chapman & Hall.

Salmon, Wesley C. 1990. "Rationality and Objectivity in Science; or Tom Kuhn Meets Tom Bayes." *Minnesota Studies in the Philosophy of Science* 14: 175–204.

Shaw, Jamie. 2022a. "On the Very Idea of Pursuitworthiness." *Studies in History and Philosophy of Science Part A* 91 (C): 103–12.

Shaw, Jamie. 2022b. "Revisiting the Basic/Applied Science Distinction: The Significance of Urgent Science for Science Funding Policy." *Journal for General Philosophy of Science / Zeitschrift Für Allgemeine Wissenschaftstheorie* 53 (4): 477–99.

Sklar, Lawrence. 1981. "Do Unborn Hypotheses Have Rights?" *Pacific Philosophical Quarterly* 62 (1): 17–29. https://doi.org/10.1111/j.1468-0114.1981.tb00039.x.

Sober, Elliott. 2008. *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press.

Solomon, Miriam. 2001. *Social Empiricism*. MIT Press.

Stanford, P. Kyle. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford University Press.

Stegenga, Jacob. 2024. "Fast Science." *The British Journal for the Philosophy of Science*, ahead of print, January 26. https://doi.org/10.1086/729617.

Stengers, Isabelle. 2018. *Another Science Is Possible: A Manifesto for Slow Science*. Polity.

The RECOVERY Collaborative Group. 2020. “Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19.” *New England Journal of Medicine* 383 (21): 2030–40. https://doi.org/10.1056/NEJMoa2022926.

Tilling, Robert I., Lyn J. Topinka, and Donald A. Swanson. 1990. *Eruptions of Mount St. Helens : Past, Present, and Future*. U.S. Geological Survey.

Warburton, D., and F. C. Fraser. 1964. “Spontaneous Abortion Risks in Man: Data from Reproductive Histories Collected in a Medical Genetics Unit.” *American Journal of Human Genetics* 16 (1): 1–25.

Wray, K. Brad. 2021. “Feyerabend’s Theoretical Pluralism: An Investigation of the Epistemic Value of False Theories.” In *Interpreting Feyerabend: Critical Essays*, edited by Jamie Shaw and Karim Bschir. Cambridge University Press. https://doi.org/10.1017/9781108575102.005.