Torsten Wilholt[*]

Epistemic Trust in Science

Manuscript version. The paper is forthcoming in *The British Journal for the Philosophy of Science* and will be available on bjps.oxfordjournals.org .

Abstract

Epistemic trust is crucial for science. The paper aims to identify the kinds of assumptions that are involved in epistemic trust as it is required for the successful operation of science as a collective epistemic enterprise. The relevant kind of reliance should involve working from the assumption that the epistemic endeavors of others are appropriately geared towards the truth, but the exact content of this assumption is more difficult to analyze than it might appear. The root of the problem is that methodological decisions in science typically involve a complex trade-off between the reliability of positive results, the reliability of negative results and the investigation's power (the rate at which it delivers definitive results). Which balance between these is the "correct" one can only be determined in light of an evaluation of the consequences of all the different possible outcomes of the inquiry. What it means for the investigation to be "appropriately geared towards the truth" thus depends on certain value judgments. I conclude that in the optimal case, trusting someone in her capacity as an information provider also involves a reliance on her having the right attitude towards the possible consequences of her epistemic work.

[*] Institut für Philosophie, Leibniz Universität Hannover, Im Moore 21, 30167 Hannover, Germany; Email: torsten.wilholt@philos.uni-hannover.de .

Keywords

1. Introduction

The expressions "trust" and, more specifically, "epistemic trust" are often used in the social epistemology of the sciences. Epistemic trust, it seems, is a particular sort of trust. We may trust the scientists who conduct research on venomous snakes to keep the objects of their study locked away safely, but that is not what we mean by epistemic trust. To invest epistemic trust in someone is to trust her in her capacity as provider of information. What exactly this involves will be the subject of this paper.

The paper's title of course contains an ambiguity. On the one hand, the division of cognitive labor within the sciences requires scientists to regularly invest trust in each other's work. On the other hand, scientists and scientific institutions are also trusted providers of information for non-scientists. Policy-makers, legislators, investors and activists as well as "ordinary people" in their capacities as citizens or consumers frequently rely on the results of science, trusting that these will help them make well-informed decisions. I will begin by focusing on the first kind of epistemic trust in science—the trust within science that keeps the fabric of the social enterprise of research together. However, I will return to the kind of trust that science receives "from without" at a later point.

In my approach to the phenomenon at issue, I will make use of a distinction between trust and reliance that is common in ethics. As Annette Baier (1986) has pointed out, the expressions are not synonymous. While trusting someone to do something is necessarily a way of relying on her to do it, the reverse is not true—trusting is a special kind of relying. Take the example of an undercover police officer following a kidnapper after the handover

2

of the ransom, relying on the criminal to lead him to the place where the hostages are hidden. In this case, it would be an unsuitable use of the word to say that the officer *trusts* the kidnapper to lead him to the hostages. To rely on person *P* to do *A* means to work the assumption that *P* does *A* into one's plans and decisions. To trust *P* to do *A* involves more. In Baier's own analysis, trusting is the kind of relying that makes the trusting person dependent on *P*'s good will. In trusting you to take care of my belongings, for example, I do not work merely from the assumption that you will do so, but rather specifically from the assumption that you will do so because you are somehow well-disposed towards me. (While Baier deserves most of the credit for alerting philosophers to the difference between trust and reliance, it should be emphasized that my use of the distinction in this paper is strongly influenced by Richard Holton's analysis.)[1]

I will start from the insight that reliance is a more general phenomenon than trust and first attempt an account of epistemic reliance in the sciences. Only later will I try to address the question whether mere epistemic reliance is good enough to keep the social fabric of science together, or whether this requires genuine epistemic trust.

2. Epistemic reliance within the sciences

What exactly does the kind of epistemic reliance that facilitates the collective cognitive enterprise called science consist in? In other words, what exactly are the assumptions about

---

[1] In particular, I follow Holton (1994) in not requiring that the relying person believe that *P* does *A*. In my attempted escape from the lion's den I may rely on the rotten rope not to break, even if I cannot summon up the confidence to believe that it will not. The same holds for trusting. This is why I speak of "working something into one's plans" (a phrase adapted from Holton) rather than, e.g., "accepting something". Another advantage of not treating epistemic reliance as a decision to believe or "accept" a reported proposition is that it avoids the problems of epistemic voluntarism (on which see Williams 1973).

their fellow researchers' performance that scientists need to work into their plans and decisions?

The assumption that other scientists speak truly or even knowledgeably when they report their results would certainly be too strong. The way that scientists initially set up their research projects may often look as if they were working from the assumption that the results they are taking from the published research literature are true, but they are usually prepared to take a step back and challenge and test the results of others in the face of unexpected experimental results. If my whole behavior in the course of an investigation was premised on the assumption that p is true, calling p into question could under no circumstances be a result or a step within the investigation. While this may arguably be the case for some core assumptions—the ones that play a constitutive role for the paradigm or research program to which the investigation belongs—it is not the rule. The assumptions about the information that scientists rely on which they work into their plans and decisions are usually weaker. And this is certainly how it should be in a well-designed collective cognitive endeavor, so that one person's mistake can be corrected by another's scrutiny. This is not to deny that reliance in the sciences has to do with some sort of confidence that epistemic endeavors of others are appropriately geared towards the truth. (To the contrary, my analysis of epistemic reliance will incorporate this assumption and is thus compatible with the veritistic program in social epistemology.[2]) But being geared towards the truth is certainly not the same as having true results and turns out, as I shall argue, to be surprisingly difficult to explicate in a precise manner.

---

[2] On which see Goldman 1992, 192-197 and 1999, chapters 1-3.

Working from the assumption that other scientists are sincere when they communicate their findings would be too weak as a basis for epistemic reliance. While sincerity is *part o*f what scientists assume about each other, it would not suffice to explain how the cognitive division of labor works. In relying on the work of others, scientists assume more than just that what the others say they have found is what they actually believe to have found. They also assume that their colleagues "know what they are doing"—that they have chosen the right methods for the problem, that they have employed them skillfully and carefully, that they have drawn their conclusions with due attention to all the observations they were able to make in the course of inquiry, and so on.

An appealing way of summing up all these assumptions might be seen in the concept of reliability. After all, it sounds reassuringly tautological to state that epistemic reliance in the sciences consists in working from the assumption that fellow scientists are reliable information providers. Taking my clue from the reliabilist approach in epistemology pioneered by Alvin Goldman (1979), I assume that the relevant reliability assumption is best understood as a supposition about the objective conditional probability that S is true given that one of my peers reports S as a result. In relying on the results of others, scientists might be considered to work from the assumption that this probability is "high enough".

This way of analyzing epistemic reliance almost automatically brings up the question: How high is high enough? As reliability is a matter of degrees, reliance has not been fully explicated as working from an assumption about reliability unless we can say something about the relevant degree of reliability. This specification could take one of two forms. There could be one assumed standard of reliability for all science, i.e. one single threshold level x, such that each and every case of epistemic reliance in the sciences consists in working from the assumption that a peer's report is reliable to at least degree x. Or the

degree of reliability presupposed in each act of reliance could vary from case to case. In this case, the analysis should have something to say about the rationale according to which it varies.

To address the first option, let us consider the question: Could there be a single degree of reliability for communicated scientific results that is binding for all of science (or at least for a certain scientific discipline or specialty) and due to the objective requirements of rationality as they present themselves to the individual researcher? The question needs to concern us if we aspire to an analysis of reliance that also leaves some hope for the widespread epistemic reliance within the sciences to be by and large justified. For this to be the case, the degree of reliability presupposed in epistemic reliance should by and large correspond to an actual standard of reliability for communicated scientific results. However, philosophers of science have long known an argument that undermines hopes for an objectively binding standard of reliability in this sense. The argument was first considered by statisticians (Wald 1942, pp. 40-41; Churchman 1948, ch. 15) and brought to the attention of philosophers by Richard Rudner (1953); discussion about it has recently been revived by Heather Douglas (2000) and others. For the purposes of this paper, I will interpret it as an argument to the effect that from an individualistic perspective, there is no non-arbitrary way to determine a correct level of reliability that applies across a whole range of investigations.[3] The "individualistic perspective" in question means that for the moment I am abstracting from the need to coordinate cognitive efforts amongst the

---

[3] The argument has initially been proposed in support of the claim that accepting hypotheses in science necessarily presupposes value judgments, but it *ipso facto* undermines prospects for a value-independent reliability standard. I have presented a detailed analysis and assessment of the argument in Wilholt 2009.

participants in a collective epistemic enterprise. Incorporating the need for coordination will make a difference, to which I will turn in section 3.

The point of departure for the argument under consideration is the essential characteristic of all empirical investigations that the evidence never provides conclusive support for any hypothesis. In order to communicate something as the result of an investigation, we have to select some level of empirical confirmation and declare it sufficient for the investigation at issue. Inductive logic and the science of statistics provide no guidance as to the choice of this level. The only facts that are rationally relevant for the question are our evaluations of the possible outcomes of the investigation—in particular, how good the consequences of a correct outcome and how bad the consequences of an erroneous one would be. In Rudner's words: "How sure we need to be before we accept a hypothesis will depend on how serious a mistake would be." (1953, p. 2)

How deep this point goes and how exactly it affects questions of reliability can be illustrated with the aid of the following Bayesian model. In this model, the only two options under a researcher's consideration are to either communicate S as a result of her empirical investigation or not to communicate any result (neither S nor non-S). Let c, e, i and j represent the researcher's evaluations of the four possible outcomes of her decision in terms of utilities:

	S is in fact true	S is in fact false
Communicate S	c	e
No result	i	j

In terms of constraints on the researcher's utilities we are only presupposing that a true result be preferred over ignorance ($c > i$) and ignorance over an error ($j > e$). The expected utilities of the two options are plotted against the probability of S in figure 1.
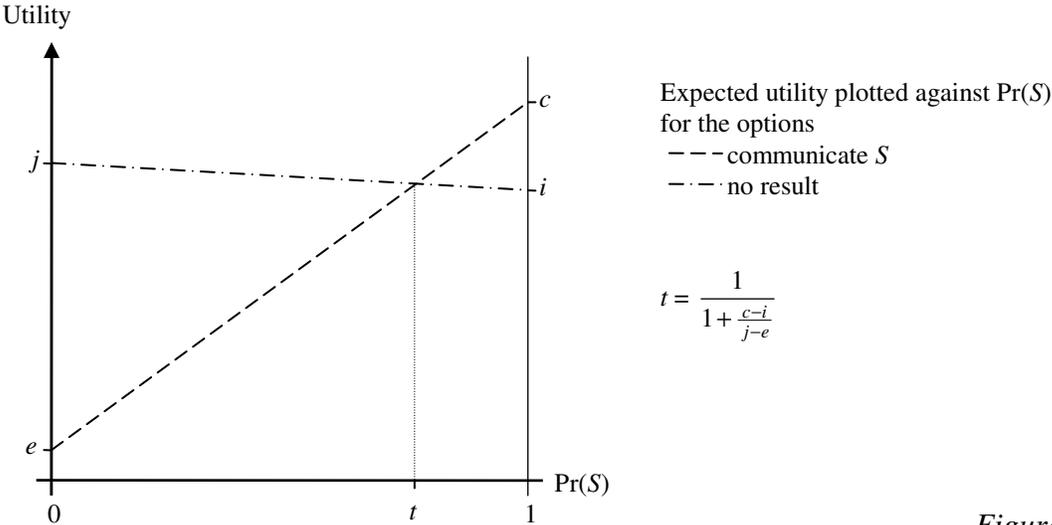

Utility

c

j

i

Expected utility plotted against Pr(S)
for the options
$---$ communicate S
$-\cdot-$ no result

$$t = \frac{1}{1 + \frac{c-i}{j-e}}$$

e

Pr(S)

0 t 1

Figure 1

The figure illustrates that communicating S as a result of the investigation maximizes expected utility if and only if the probability of S is t or higher, where $t = 1/(1 + (c - i)/(j - e))$. A rational agent who conducts a finite investigation into the question whether S, collecting evidence and updating her subjective probabilities accordingly along the way, would therefore communicate S as a result of the investigation if and only if she ends up assessing the probability Pr(S) with a value of t or more.

Let us reflect how this may be related to the (objective) reliability of a positive result of the investigation. In order to infer anything at all about this reliability, we will have to assume that the agent's subjective probability assignment at the end of the investigation constitutes a good approximation of the objective probability of S given the kind of

8

evidence that was collected.[4] But even under that assumption, it would be wrong to conclude that the reliability of a positive result would be t. If the "investigation" is of a simple sort, where strong evidence is easy to come by (such as counting the books on a shelf), the subjective probability assigned by the agent at the end of the investigation is likely to be very close to either 1 or 0. Positive results under such circumstances can thus only be said to be reliable *at least* to degree t—the actual reliability is likely to be much higher.

However, the reliability can be expected to be close to t in cases where strong evidence is hard to come by and resources for collecting evidence are scarce. We can include the aspect of resources into our considerations indirectly, by supposing that the researcher collects her evidence stepwise. After each step, she decides whether or not to communicate S as a result of her investigation in accordance with the decision matrix described. In the event that she settles on "no result", she then makes another decision (not modeled by our matrix) whether or not to invest additional resources to investigate the matter further. Assuming that she starts from ignorance ($\Pr(S) \approx 0.5$) in the first step and that strong evidence is difficult to obtain, such that the subjective probability $\Pr(S)$ changes only slowly over the course of each such iteration, a communicated result will typically reflect the fact that the researcher has in the last step arrived at a subjective probability just over t.

What this shows is that, if an investigation is difficult in the sense just described, the reliability of its result should be expected to be highly sensitive to whichever utilities are

---

[4] This assumption is perhaps best understood as the claim that in a long series of investigations into questions $S_1, \ldots, S_n$ which are similar to the present investigation into S with regard to the evidence collected (and the kind of question asked), the results $S_1, \ldots, S_n$ tend to be true at a relative frequency corresponding to $\Pr(S)$.

assigned to its possible outcomes. The one crucial measure, it turns out, is the ratio between the intervals $c - i$ and $j - e$. Interestingly, and just as an aside, this allows us to observe that the high degree of reliability which we usually expect from scientific investigations depends on an implicit utility structure in which the extent to which a false result is bad news (as compared to no result) is much greater than the extent to which a true result is good news (as compared to no result).

It might be objected that in practice, researchers have more finely calibrated options at their disposal, such as communicating something as a result that is "strongly suggested by the data," or other ways of qualifying their communication of a result. But including these additional options would not change the picture in principle. Each way of qualifying the communication of a result just adds another row to the decision matrix. Researchers would still have to decide whether their result is well-confirmed enough to count as "strongly suggested by the data", for example, and this can only be done by taking the utilities into account.[5]

Note that the decision whether or not (and together with which qualifying caveats) to communicate S as a result at the end of an investigation is of course not the only choice through which the researcher influences the ultimate reliability of her output. In deciding whether or not to discard a set of measurements from a particular run of an experiment, or how to analyze and interpret a set of data, even in settling the details of the design for a new experiment or study she faces a similar kind of choice. In general, whenever a researcher has to make a yes-or-no decision that affects the further course of inquiry and influences its

---

[5] This objection was pressed against me by an anonymous referee for this journal, who was also kind enough to suggest another response to it: Another reason why the option of adding qualifications and caveats does not change the picture very much is that they are typically not listened to.

inherent inductive risk (i.e., the probability of ending up with a false result), the rational way to make that decision from a purely individualistic perspective should reflect the utilities c, e, i and j.[6] If the investigation is a difficult one and the researcher has to economize on her resources, she would ideally combine her methodological choices in such a way as to aim at a reliability of just over t, or, to put it differently, to limit the inductive risk (of a false positive) to just under $1 - t$.

Many of the methodological decisions in question are deeply embedded within the research process, and are often not explicitly documented and communicated in each individual case. In an example described by Douglas (2000), it is the examination of individual slices of rat livers for signs of tumors, and in particular the fashion in which researchers deal with the many borderline cases in assessing them, that the inductive risk of a toxicological animal experiment critically depends on. Countless small yes-or-no decisions may have to be made that are sensitive to an assessment of how sure we need to be in order to report a certain result at the end of our inquiry. The fact that most of these are typically not communicated is not just a contingency of current publication practices. Some methodological decisions, such as those described by Douglas, may involve tacit forms of judgment that would be impossible to fully document and communicate even if time and space played no role. Moreover, "mere" restrictions of time and space should always be taken seriously when it comes to communication. Full documentation of all methodological decisions would only make a difference if it was also fully absorbed by the recipients and users of the results, which is a practical impossibility. In order to make the division of cognitive labor practically feasible, publications must compress information.

---

[6] The term "inductive risk" was introduced in this sense by Hempel (1965, 91-92).

Douglas stresses the point that the values which determine the acceptable inductive risk in each case must include social, political and moral values. That conclusion is plausible if one regards the utilities c, e, i and j as primarily reflecting the consequences that a published research result leads to via its applications.[7] But it is not necessary to argue this point in order to support the conclusion that interests me here, which is the insight that the decisive ratio $(c – i)/(j – e)$ does not have one and the same value for all propositions that are investigated by a scientific specialty. This insight follows no matter whether the utilities considered are the researcher's own personal utilities, or ones that reflect a whole research community's collective values, or even an abstracted sort of "cognitive utilities" that express only the epistemic (as opposed to political and moral) concerns of the scientific community.[8] Even if we consider the latter kind of utilities, ignoring for the moment the difficulties of separating cognitive from non-cognitive concerns in science (cf. Longino 1996), the cognitive utility of validating a true proposition depends on a number of features specific to that proposition. Science cannot pursue each truth with the same eagerness (cf. Kitcher 1993, ch. 4, 2001, ch. 6), and so utilities that express the cognitive relevance of validating a proposition must differ from case to case. A proposition's value for the systematic organization of our beliefs, its explanatory power and its potential to lead to

---

[7] Against this understanding of the situation of methodological choice, Richard Jeffrey (1956) had argued in response to Rudner that scientific results often have either more than one or no foreseeable applications and that therefore, accepting or rejecting hypotheses cannot belong to the scientist's job description. Instead, Jeffrey suggests that the scientist should just attribute probabilities to hypotheses and leave the decision-making to politicians and other extra-scientific practitioners. However, as the preceding observations have made clear, some yes-or-no decisions are inevitably involved in the practice of science, and the scientists have to make them on the basis of some utility function or other.

[8] On cognitive utilities see Maher (1993).

more fruitful research in the future are all features that vary among propositions and lead to differing utilities even if only "purely epistemic" concerns are taken into account. Nor can it be expected that the "costs" of false positive errors ($j - e$) will magically stay in fixed proportion to the benefits of the respective true results ($c - i$) and thus make for a constant value of t. What is true for cognitive utilities, that their structure must be expected to vary from proposition to proposition, *a fortiori* holds for other kinds of collective or individual utilities which only add more factors that can cause variation. The fact that in actual practice various scientific methods differ widely in their reliability should be obvious enough. The above considerations show that, as long as we restrict ourselves to the perspective of the individual researcher and abstract from the need to coordinate epistemic efforts among individual agents, one universal standard of reliability for a whole scientific specialty or discipline (let alone all of science) would not even make sense as an ideal, because such a standard would not do justice to the differing utilities of different scientific results.

The same considerations also reveal a deep problem for the other possibility to consider, that the degree of reliance could itself vary from one case to the next, always presupposing a degree of reliability that is appropriate for the respective case. As we have seen, the methodological decisions that affect reliability often occur deep within the research process. They are often not explicitly communicated along with the results. This seems to indicate that it should be very difficult to judge the reliability that a given research result even aims at. So if epistemic reliance consisted in working from the assumption that reported results of others are reliable to a degree which can vary from case to case, that

would give rise to serious difficulties in explaining how researchers choose the appropriate degree in each individual case of reliance.[9]

The problem is complicated by the fact that in many situations, assessing the quality of results with regard to one particular research question requires the consideration of more than just one type of inductive risk. A scientific investigation into the question whether S often also gives rise to the possibility of finding non-S as a result. There are therefore two more utilities (in addition to the ones we have already considered) in light of which methodological decisions will be made and evaluated:

	S is in fact true	S is in fact false
Communicate S	c	e
No result	i	j
Communicate non-S	f	d

These additional utilities will directly affect the projected reliability of negative results, through their part in determining the acceptable inductive risk u of false negative errors (see figure 2).[10] (I will reserve the term "false negative" for investigations that result in non-S as a communicated result where S is in fact true.[11])

---

[9] Kitcher's (1993, ch. 8) influential treatment of scientists' reliance on each others' results focuses on the important problem of how to assess another researcher's *competence* to achieve reliable results. The problem I present here, of identifying the degree of reliability that other researchers even aim at, should be considered as an additional one that adds to the difficulties addressed by Kitcher.

[10] In analogy to the restrictions already introduced, I am going to assume that $d > j$ and $f < i$. I am also going to presuppose (here and in the rest of the paper) that $(d - j)/(j - e) < (i - f)/(c - i)$, meaning intuitively that the utilities are distributed in such a way that the expected utility function for the "no result" option passes above the point where the other two intersect. (This effectively disallows utility distributions which would motivate a rational researcher to design an investigation that always comes up with either a positive or a negative result
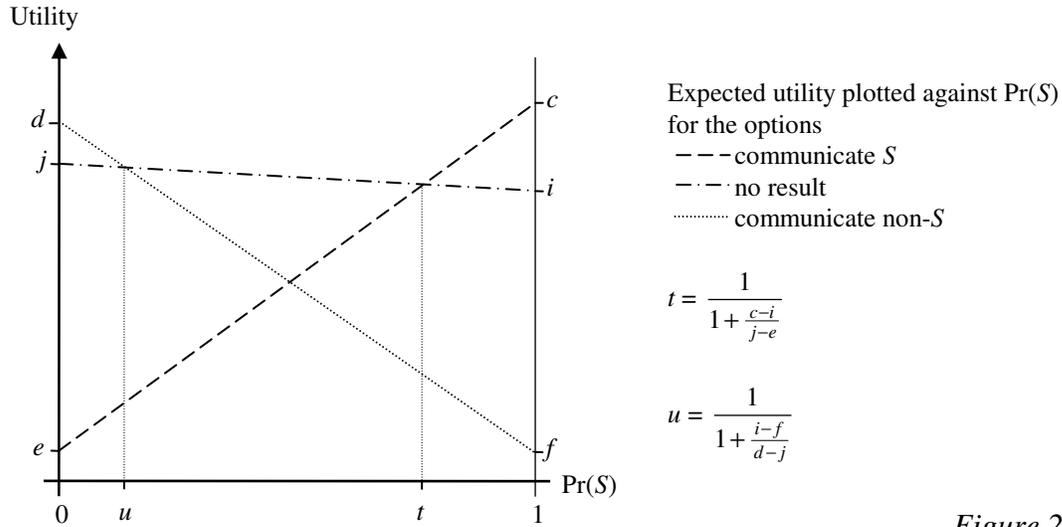
Utility

d

j

c

i

e

f

Pr(S)

0   u t 1

Expected utility plotted against Pr(S)
for the options
--- communicate S
--- no result
.... communicate non-S

$$t = \frac{1}{1 + \frac{c-i}{j-e}}$$

$$u = \frac{1}{1 + \frac{i-f}{d-j}}$$

Figure 2

In practice, one and the same methodological decision will very often affect the probability

of arriving at a positive result *and* the likelihood of ending up with a negative one.

Methodological decisions therefore typically involve a complex trade-off between the three

risks of committing a false positive error, committing a false negative error and ending up

without any result at all. In scientific practice, the reliabilities of positive and negative

results are thus likely to be inter-dependent and affected by evaluations corresponding to all

six of the utilities represented in figure 2.

3. Methodological conventionalism

We have seen that if epistemic reliance entails working from an assumption of reliability,

there is a general difficulty involved in identifying the appropriate degree of reliability that

may reasonably be assumed in each case—at least as long as methodological decisions are

---

no matter how inconclusive the evidence.) Nothing of substance depends on the latter presupposition, except

that to do without it would require us to make some annoying case differentiations.

[11] I thereby diverge from an alternative broader use of "false negative" that covers all cases of failing to report

S where *S* is true, including "no result" cases.

regarded as a matter of individual rationality. This is a potential threat to the efficient division of cognitive labor in the sciences. When reliability is underestimated in acts of reliance, results of others will be called into doubt too early, leading to an unnecessary duplication of scientific work. On the other hand, in cases of excessive reliance where reliability is overestimated, researchers will hold on to the results of others more persistently than is warranted, which will in some cases lead them into dead ends that could have been avoided.

The problem of epistemic reliance is thus a problem of coordination. The difficulty is not that epistemic reliance in the sciences requires a particular, definitive degree of reliability either on the side of those who rely or on the side of those who provide the results that others rely upon. Rather, the social epistemology of the sciences requires that the reliability that producers of information aim at and the reliability that is presupposed in acts of epistemic reliance be suitably adapted to one another.

In very general terms, the appropriate practical solution to a problem of coordination is a convention (Lewis 1969, ch. 1). In pursuit of this basic idea, I will now try to show how the social problem of epistemic reliance is in practice at least partly solved by means of conventional methodological standards.

That there are some conventional standards in place that have a direct bearing on the problems at issue is easily illustrated. The most obvious example is perhaps the convention (which can be found in a whole range of scientific specialties) that .05 is the highest reasonable significance level in significance testing. If the stipulation was .04 instead, "significant" results would on the average be more reliable, but also more difficult to obtain. If it was .06, it would be the other way round. One would be hard pressed to argue

that .05 represents just the "best" choice for facilitating a socially organized cognitive endeavor. What is crucially important, instead, is that there is one level that is widely recognized and that thereby helps researchers to gauge their reliance on reports that communicate the finding of a "statistically significant effect".

However, the range of methodological standards that can be regarded as conventions in this fashion is much broader. As was mentioned before, all kinds of decisions about experimental setup, data analysis and interpretation affect the distribution of inductive risks. Many of these are constrained by methodological rules that are shared among the members of a research community: rules about what kind of controls a study should include, rules about what kind of precautions should be taken to causally isolate an experiment, rules about the conditions under which a set of data may be discarded, and so on.

It may be thought that such rules cannot be regarded as conventional stipulations because they serve to increase the reliability of experiments and studies and are therefore not arbitrary (i.e., they are not really solutions to pure coordination problems). But the above considerations show that increasing the reliability of a method cannot be the one unconditionally desirable methodological aim—increasing the reliability of one kind of result will always come at a cost, usually the cost of an increased occurrence of investigations with no result (and possibly an increased risk of false negative errors too). Methodological standards must always strike a balance between reliability and power, where "power" signifies the rate at which a scientific investigation delivers definitive results.[12] Which precise balance is struck will always be arbitrary to at least some extent. What is decisive however is that the methodological standards impose certain constraints

---

[12] This use of the term "power" is adapted from Goldman (1992, 195), where it is introduced to signify the rate at which a practice delivers true results (thus combining reliability and power in my sense).

on the balance, and that these constraints and the ways in which they are imposed by the standards are common knowledge within the research community.[13]

Such conventional methodological standards facilitate epistemic reliance within science. A researcher learning about the result that a fellow scientist has reportedly achieved with the aid of a certain method will know the conventional standards pertinent to that method and have a sense of the balance between reliability and power that they strike. This knowledge will give her guidance—for example with regard to the question as to how long she should hold on to the reported result in face of recalcitrant experimental results of her own.

How would we best describe the kind of assumption about other scientists' results that researchers build into their plans and decisions in acts of epistemic reliance of the sort that is fostered by conventional standards? I submit that we should not regard it as a straightforward assumption about reliability. First of all, many standards implicitly set limits to a range within which the balance between reliability and power may be struck rather than directly determining one definite balance. Secondly, not all methodological decisions are standardized, nor can they be. The ever-developing practices of science will always force their practitioners to make choices (such as whether to try out some new experimental technique or not) that cannot be constrained by conventional standards (yet).[14] Both points indicate that the assumption at the focus of an act of reliance of the sort

---

[13] Common knowledge in the relevant sense requires that everybody knows and that everybody knows that everybody knows, such that the system of mutual expectations that typically characterizes convention can arise. Cf. Lewis 1969, 52-60.

[14] If more reasons are needed, the relevance of skills and tacit knowledge for scientific practice could also be appealed to in order to support the claim that not everything in science can be standardized.

considered here is one that offers the relying researcher guidance on her assessment of the pertinent reliabilities rather than itself being an assumption about reliability. More precisely, it seems that what the relying scientist works into her plans and decisions is the assumption that the results she is relying upon were arrived at by means of professional methods suitably employed (where the precise meanings of these terms are determined by the conventional standards of the pertinent research community). In other words, the phenomenon at issue is reliance upon widespread observance of the rules of the trade. It can very plausibly be expected to provide a basic structure of epistemic reliance that plays a part in keeping the social-epistemic fabric of science together.[15]

4. Trust in science

But is the kind of reliance which is fostered by conventional methodological standards sufficient to facilitate division of cognitive labor as practiced in the sciences? We saw that the common knowledge about standards provides guidance for assessing the reliability of a result. But if standards often leave some maneuvering space for methodological decisions, and if some decisions are not standardized at all, how are researchers supposed to fill the gaps in their knowledge about how exactly the reported result came about in order to get an idea of its intended reliability? Recall that methodological decisions often occur deeply embedded in the research process and are often not made explicit in the published research literature (nor *could* they be, because the efficient division of cognitive labor does not

---

[15] My description of methodological conventionalism is intended to preserve the spirit of an idea that Isaac Levi succinctly articulated in response to Rudner's problem: The values that determine the acceptable levels of inductive risk, Levi thought, might be considered part of the set of normative principles that a scientist commits herself to when she commits herself to "certain 'scientific' standards of inference". (Levi 1960, 356)

permit arbitrary amounts of effort to be spent on communicating every last detail of the methodology).

If not all the relevant methodological decisions are themselves transparent, then the only way of factoring in their import in a non-arbitrary way that remains open to those relying on a research result is to work from an estimation of the utilities in light of which the decisions are likely to have been made. In some contexts, for example, researchers might expect each other to make all their methodological decisions in a "disinterested" fashion, meaning that they ought to value all correct outcomes equally high ($c = d$) and all errors equally low ($e = f$, $e \ll c$) and evaluate all states of ignorance on an even level somewhere in between ($i = j$, $e < i < c$).[16] If all methodological choices are made rationally in accordance with these utilities (and are based on realistic assessments of the probabilities of S in light of various possible bodies of evidence) then one should expect the reliabilities of both positive and negative results to be equally high, because the investigation would be geared towards reporting S if and only if $\Pr(S) > t$ and reporting non-S if and only if $\Pr(\neg S) > 1 - u$, and under the restrictions described, $t = 1 - u$ (see figure 3).[17]

---

[16] This is in effect the utility structure that Levi (1962, pp. 55-56) ascribes to a person seeking the truth and nothing but the truth.

[17] The restrictions in question are not necessary for effective disinterestedness in the sense of $t = 1 - u$. Generally, $t = 1 - u$ will hold as long as $(i - f)/(d - j) = (j - e)/(c - i)$, which would also permit more "unequal" utility distributions such as $d = 7$; $j = 6$; $e = 0$; $c = 6$; $i = 4$; $f = 1$.
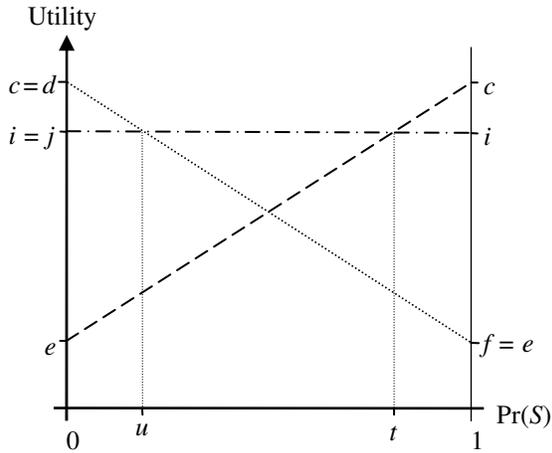
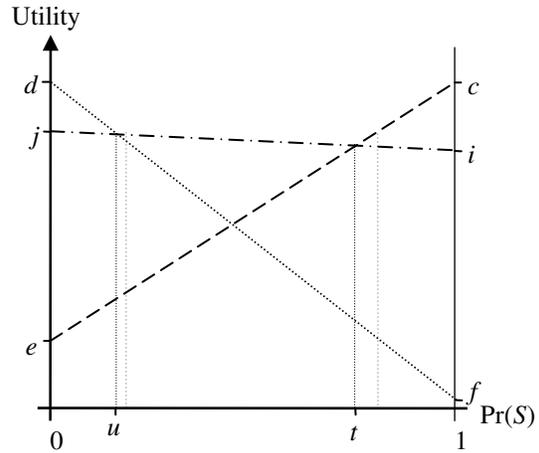*Figure 3*   *Figure 4*

This would mean that disinterested methodological decisions could not involve any trade-off between the risk of false positive errors and the risk of false negative errors. It would, however, not prevent trade-offs against the probability of ending up without any result and would in that sense not determine how the balance between reliability and power is struck. In order to assess this additional dimension, one would have to make further assumptions about where exactly between c and e the relative value of ignorance i has been set. This, more precisely the ratio $(i - e)/(c - e)$, might be called the degree of caution of the respective investigation.[18]

Such a simple kind of "disinterested" utility structure will not always be the one that scientists expect each other to apply to their methodological decisions. Consider, for example, a toxicological investigation into the question of whether substance X poses a health risk to infants when used as a softening agent in plastic baby bottles. In such an investigation, given the particularly dreadful results that would ensue if substance X in fact

---

[18] Cf. Levi 1962, 56-57, who is well aware of this problem and introduces the term "degree of caution" for a value that is equivalent to our $(i - e)/(c - e)$.

represented a major health risk to babies but our investigation failed to identify this risk, one may reasonably expect the utilities for *i* and *f* to be significantly reduced as compared to a purely academic investigation. The difference might be illustrated by the shift from figure 3 to figure 4, which also makes intuitively clear that the specific circumstances would justify relaxing the projected reliability *r* for positive results but tightening the standard $1 - u$ for negative results.[19]

To be sure, I do not mean to suggest that considerations about the consequences of their results justify scientists to take any methodological steps in defiance of existing conventional methodological standards. The above considerations are intended to pertain only to the kinds of methodological choices that are left open by the existing standards. Relying on blunt standards is arguably easier and less error-prone than relying on a shared sense of the appropriate value judgments and the latter may therefore provide only the second best solution to the problem of coordination. However, it is not practically feasible to standardize a complex real world practice down to the last detail, let alone a practice as innovative and ever-changing as science.

With regard to the decisions that take place within the leeway that the existing standards thus always have to leave open, disinterestedness may sometimes be a feature of the utility structure that scientists expect to underlie each others' choices, and sometimes not. Even where it is, this does not settle the question before an additional assumption about

---

[19] The choice of example should not mislead us to believe that disinterestedness will be the rule as long as no "external" factors come into play. Also purely "internal" considerations can lead to an asymmetrical distribution of utilities. For the systematic organization of our knowledge, for example, the value of confirming that a certain effect or phenomenon exists might be much higher than the value of ascertaining that it does not exist.

the degree of caution has been made. All this supports at least the following observation: Epistemic reliance in the sciences and thus the division of cognitive labor can be expected to work even more efficiently than on the basis of conventions alone if the scientists within a research community share certain value judgments. The value judgments in question are of course those that are reflected in the sets of utilities ascribed to the possible outcomes of the investigation. As we have seen, these are in effect value judgments about the benefits of correct results relative to the costs of incorrect ones.

To be more precise, it is not strictly necessary that the researchers agree with each other in their honest personal value judgments. It suffices if the evaluations of possible outcomes that the information users assume the information producers to have taken as a basis for their methodological decisions and the evaluations that the latter have actually used are (approximately) the same. This presupposes only a shared sense of what the right evaluations would be.

In this way there arises the possibility of an enhanced kind of epistemic reliance—reliance based on the presumption of shared ideas about the values of true results and the dangers inherent in errors. This kind of reliance presupposes much more than just that other scientists work dependably and professionally in keeping with the rules of the trade. It presupposes that they have the right attitude towards what they are doing—an attitude whose absence might be considered not just regrettable but to a certain degree blameworthy. I therefore suggest that this enhanced kind of epistemic reliance would be

truly deserving of the name of epistemic trust. It is, after all, a kind of relying that makes me dependent on another's working from the right value judgments.[20]

The fact that it is theoretically possible to work assumptions about other people's value judgments into my plans and decisions without actually sharing these value judgments might be taken to indicate that genuine epistemic trust in the above sense is not really required for the kind of coordination at issue. In theory, the coordination could even be successful on the basis of a set of value judgments that *no-one* really made, but that everyone just assumed to be the kinds of value judgments that are commonly expected and applied. But this way of achieving coordination, while theoretically possible, would in all likelihood suffer from a lack of stability. The coordination would break down as soon as it was discovered that the value-judgments were not widely held. Genuine epistemic trust is what makes the stable coordination of epistemic efforts practically feasible.[21]

---

[20] Obviously this does not exactly coincide with Baier's (1986) initial analysis of trusting as the kind of relying that makes me dependent on another's good will. But her conception (which is built around the good will of one trustee toward one trustor) cannot be transferred unchanged to the case of a collective enterprise bound together by trust, which is our present concern. In such an enterprise, it should also be possible that the "good will" (or a comparable kind of disposition) is toward the enterprise itself rather than toward its participators. An analysis of the difference between trust and mere reliance that may lend itself more easily to the collective case is Carolyn Mcleod's (2000) suggestion that trusting is the kind of reliance that involves optimism about the trusted person's moral integrity.

[21] It might be objected that even if I *do* share someone's value judgments I can rely on them in a sober and calculating mode, without feeling let down if they fail to do as I expected. Would that not mean that my reliance need not amount to genuine trust? It would, if the affective response of feeling let down was a constitutive element of trust. I find it more convincing to see the decisive element in the particular kind of motivation that we assume to be at work in a person we trust – the kind that is rooted in her good will, or in her moral integrity. If such trust turns out to have been misplaced, to feel let down is a typical and perhaps

This assessment also fits well with another characteristic that has been used to set trust apart from mere reliance, namely its interactive component. When I trust, Philip Pettit has noted, I typically believe that making my reliance manifest to the trustee will strengthen her reasons to do that which I rely on her to do (1995, p. 206). Shared ideas about what the right evaluations of an inquiry's outcomes would be, if they exist, can be expected to display exactly this kind of interactive element. The very fact that my peers rely on me to base my decisions on certain valuations would, once it becomes known to me, strengthen my reasons to do so.

The question remains: Does the collective enterprise of science need the kind of coordination of value judgments that is facilitated by genuine epistemic trust to bind it together, or is mere reliance enough? One reason to think that genuine trust is required I have already mentioned. Not everything can be standardized, and thus the mechanism of reliance that is supported by methodological standards leaves too many gaps to facilitate a dependable assessment of reliabilities in and of itself. Lest these be filled by mere guesswork, assumptions about outcome evaluations must enter into the picture.

But an additional reason should also be considered. It has to do with the other kind of trust in science, the trust that is invested in scientific results by those outside the scientific communities. The reliance of non-scientists upon scientific information often consists in them working from the assumption that science is a disinterested endeavor. Trade-offs between the risks of false positive and false negative errors might then be regarded as cases of bias and as a betrayal of the trust invested in science by the public. The topic is particularly sensitive if the potential real world consequences of false positive errors on the

---

natural reaction, and thereby indicative of this particular kind of reliance, but it is not required by the definition of "trust".

one hand and false negative errors on the other would have to be borne by different parties—for example by consumers on the one hand and by producers on the other, or by developers on the one hand and by the public on the other (cf. Wilholt 2006, pp. 70-1). In some cases, however, the public may expect scientists to take exceptional care to avoid *one* of the two types of error—for example if that kind of error would have catastrophic consequences. Under such circumstances, non-scientists trust scientists to base their methodology on evaluations of the possible outcomes that, rather than being "disinterested" in the sense specified above, reflect the paramount interests of society at large.

In either case, how can the scientific community do justice to the trust invested in it from without? Only by actually basing methodological decisions on the kinds of evaluations that the public expects them to be based on.[22] As science is essentially a social enterprise, this requires the evaluations in question to pervade the research community. In order to achieve this pervasion, science needs a stronger mechanism than just conventional standards. It needs the interactive and value-infused instrument of genuine epistemic trust within the scientific community in order to be able to do justice to the trust it receives from without.[23] As I assume that it is at least sometimes and in some respects desirable that

---

[22] By this I do not mean to imply that the scientific community should always evaluate the outcomes of inquiries in line with the public. If the public trusted scientists to come up with some refutation of anthropogenic climate change so that everyone can keep driving their SUVs with a calm conscience, that would not be a good reason for the climate scientists to relax their standards. Sometimes trust is better disappointed. The public may also trust a judge to send a certain man to jail, but if the man is innocent the judge had better set him free regardless.

[23] In addition, the instrument of conventional standards would also have to be calibrated to accord with the evaluations that are supposed to pervade the whole enterprise. This would, perhaps plausibly, qualify their character as being "purely conventional": While all kinds of stipulations striking all kinds of balances between

research communities strive to deserve the trust invested in them by the public,[24] this constitutes another reason why science needs to be committed to epistemic trust and not just mere reliance.

As above, my point is not that the expectations of the public give scientists an excuse to violate methodological standards. The preceding remarks are primarily directed at the kinds of decisions that are left open by the existing methodological conventions. However, they may also be of relevance for situations in which research communities settle on new standards. The distribution of inductive risks that is implicit in a new standard should correspond to the evaluations expected by the public, or else the public trust in science will be to a certain extend misplaced.

I regard the arguments in this paper as independent from and complimentary to the reasons by which John Hardwig has supported the claim that cooperation in the sciences requires "trust in the moral sense" (1991, 702). Hardwig rests his case on the premise that the practices of peer review and replicating experiments are insufficient for the effective detection of fraud and other forms of misconduct. My reliance on other scientists' good conduct can then not solely be based on my understanding that they act in their own long term self-interest (as Blais 1987 had argued). If fraud stands a good chance of passing undetected, my confidence that others nevertheless abstain from committing it constitutes

inductive risks could function as reliance-enabling and reliance-preserving, only a limited range of them would also count as trust-enabling and trust-preserving.

[24] As I have tried to argue, there is no one set of evaluations of the possible outcomes of an inquiry that is "correct" or "properly scientific". Using the ones which the public trusts science to use would therefore not interfere with the "inner logic" of the scientific enterprise. There are in fact many positive reasons for at least sometimes doing so—the most prosaic one being that science depends on the public's patronage for its very existence.

the kind of reliance that makes me dependent on another's "character", as Hardwig puts it. The arguments that I have discussed in this paper provide *additional* reasons for thinking that the collective cognitive enterprise of science has to be bound together by trust—reasons that would in my view remain convincing even if there was an effective self-regulatory mechanism for the detection of fraud and other forms of misconduct in place.[25]

5. Conclusions

I have argued that epistemic reliance in science sometimes has to take the form of working an assumption about others' evaluations of the possible outcomes of an investigation into one's plans. I have suggested calling this kind of reliance epistemic trust, because it is based on a shared sense of what the right attitude toward the aims of a collective epistemic enterprise is and on the confidence that other participants in the enterprise actually display that attitude. Perhaps nothing of substance depends on the terminological suggestion of calling this trust, although I do consider it justified by the extent to which the mutual expectations that this kind of reliance involves are likely to be normatively charged, which is a telling characteristic of trust. In Baier's words: "[T]rusting can be betrayed, or at least let down, and not just disappointed." (1986, p. 235)

---

[25] Hardwig's influential and pioneering analysis presents epistemic reliance as a reliance on others' sincerity, competence and justification (in the sense that they themselves have good reasons for the beliefs that they testify on) (1985 and esp. 1991, 699-700). I would argue that these elements alone do not suffice to describe the structure of epistemic trust in science, because they do not allow us to characterize assumptions about the balance between reliability and power and about the distribution of inductive risks. On the other hand, a notable similarity between Hardwig's case and the one presented in this paper is that both contain an appeal to the tacit dimension, i.e. to the fact that in a collective epistemic enterprise not all the details that are relevant for the acceptance of a particular piece of information can be effectively communicated to every recipient of that information (cf. esp. Hardwig 1985, 338-339).

The substantial result of these considerations is the high degree to which the phenomenon of epistemic trust is infused with value judgments. One way of putting this is to observe that the apparently simple distinction between epistemic and other kinds of trust in science with which I began may be less clear-cut than it seems. If trusting someone in her capacity as information provider involves a reliance on her having the right attitude towards the possible consequences of her work, epistemic trust is likely to be intricately interwoven with general expectations regarding the scientist's sense of responsibility. Trust in science, whether epistemic or otherwise, always involves a normative assessment of its aims and means.

This, it should be emphasized again, does not undermine the venerable supposition that truth is the aim of inquiry, or that a consensus on aiming at the truth is important for every collective cognitive enterprise. Rather than saying too much about the social enterprise of science, as some critics of truth have maintained, the preceding observations show that such suppositions say too little. The right balance between reliability and power as well as the right balance between the risks of false negatives and false positives are underdetermined by the aim of truth. An integrated collective epistemic enterprise requires the general aim of truth to be supplemented by a more specific consensus on the benefits of getting it right and the costs of getting it wrong in each specific case.[26]

---

[26] Note that in my model, the general aim of truth is incorporated in the more specific valuations in form of the constraints $f < i < c$ and $e < j < d$.

References

Baier, Annette (1986): "Trust and Antitrust", *Ethics* 96, 231-260.

Blais, Michael (1987): "Epistemic Tit for Tat", *The Journal of Philosophy* 84 (7), 363-375.

Churchman, C. West (1948): *Theory of Experimental Inference*, New York: Macmillan.

Douglas, Heather (2000): "Inductive Risk and Values in Science", *Philosophy of Science* 67, 559-579.

Goldman, Alvin I. (1979): "What is Justified Belief?" in: *Justification and Knowledge*, ed. by George Pappas, Dordrecht: Reidel, 1-23.

Goldman, Alvin I. (1992): *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: MIT Press.

Goldman, Alvin I. (1999): *Knowledge in a Social World*, Oxford: Oxford University Press.

Hardwig, John (1985): "Epistemic Dependence", *The Journal of Philosophy* 82 (7), 335-349.

Hardwig, John (1991): "The Role of Trust in Knowledge", *The Journal of Philosophy* 88 (12), 693-708.

Hempel, Carl G. (1965): "Science and Human Values", in: idem, *Aspects of Scientific Explanation*, New York: Free Press, 81-96.

Holton, Richard (1994): "Deciding to Trust, Coming to Believe", *Australasian Journal of Philosophy* 72, 63-76.

Jeffrey, Richard C. (1956): "Valuation and Acceptance of Scientific Hypotheses", *Philosophy of Science* 23 (3), 237-246.

Kitcher, Philip (1993): *The Advancement of Science*, Oxford: Oxford University Press.

Kitcher, Philip (2001): *Science, Truth and Democracy*, Oxford: Oxford University Press.

Levi, Isaac (1960): "Must the Scientist Make Value Judgments?", *The Journal of Philosophy* 57 (11), 345-357.

Levi, Isaac (1962): "On the Seriousness of Mistakes", *Philosophy of Science* 29, 47-65.

Lewis, David K. (1969): *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.

Longino, Helen E. (1996): "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy", in: *Feminism, Science, and the Philosophy of Science*, ed. by Lynn Henkinson Nelson & Jack Nelson, Dordrecht: Kluwer, 39-58.

Maher, Patrick (1993): *Betting on Theories*, Cambridge: Cambridge University Press.

McLeod, Carolyn (2000): "Our Attitude towards the Motivation of Those We Trust", *The Southern Journal of Philosophy* 28 (3), 465-479.

Pettit, Philip (1995): "The Cunning of Trust", *Philosophy and Public Affairs* 24, 202-225.

Rudner, Richard (1953): "The Scientist *qua* Scientist Makes Value Judgments", *Philosophy of Science* 20, 1-6.

Wilholt, Torsten (2006): "Design Rules: Industrial Research and Epistemic Merit", *Philosophy of Science* 73 (1), 2006, 66-89.

Wilholt, Torsten (2009): "Bias and Values in Scientific Research", *Studies in History and Philosophy of Science* 40, 92-101.

Williams, Bernard (1973): "Deciding to Believe", in: idem, *Problems of the Self*, Cambridge: Cambridge University Press, 136-151.