BELIEF-FORMING PROCESSES AND INSTRUMENT-BASED TESTIMONY

George Borg

Contents:

1. Introduction
2. Goldberg on Epistemic Extendedness, Testimony and Instrument-Based Belief
3. Appeals to expert opinion
4. The role of assurance in (scientific) testimony
5. Assurance can’t be automated
6. Instruments should be included in belief-forming processes for purposes of epistemic assessment
7. Concluding remarks
8. References

1. Introduction

Can a computer be a witness? This kind of question is raised by the introduction of sophisticated computer programs into the criminal justice system. A recent example is a homicide case in which a computer program, TrueAllele®, was used to interpret DNA in the homicide case against Michael Robinson, accused of killing two people in Duquesne, Pennsylvania in 2013. The program found that DNA recovered from the crime scene was 5.7 billion times more likely his than coincidence (Ward 2016). One of the defendant's attorneys is reported to have said that "the witness in this case is a computer." The difficulty the defense was facing is that they wanted to know how TrueAllele arrived at the results it did. With human expert witnesses, the normal procedure in such cases is cross-examination. Unfortunately, as the attorney observed, "you can't cross-examine a computer."

Though I will have more to say about this kind of case at the conclusion of this paper, the latter will be concerned not with the use of computer programs in forensic science, but with the more general phenomenon the case illustrates: that scientific work is increasingly being carried out, in part, by machines. Though this observation is hardly news within science, for the epistemology of testimony the case is interesting because the machine was used to replace the scientist, who in this case would have been a forensic DNA analyst, as a source of testimony-based belief. Increasingly, computers are being called upon to carry out the functions of human scientists in contexts where an appeal to scientific opinion is being made. The complexity of contemporary societies as well as the efficient satisfaction of the corresponding complex social needs depend on a division of cognitive labor among experts.[1] Non-scientists are heavily dependent on scientists for the production and defense of claims

[1] See Pinto (2016) for a recent review of the philosophical literature on the division of cognitive labor. Though there are, of course, many different kinds of expertise, in this paper I will be exclusively concerned with scientific expertise. I will therefore use 'expert' and 'scientist' interchangeably.

affecting collective and individual action. Familiar examples are the roles of economics or climate science in debates concerning economic and environmental policies, or the role of the biomedical sciences in medical practice.

Testimony is one of the main means by which scientific practice is articulated with other social practices. In the ordinary, anthropocentric case, the human scientist is readily attributed epistemic responsibility for her testimony. This responsibility consists in satisfying standards of reliable testimony, which I will discuss later in the paper. Epistemic responsibility is shared by the audience, which can choose to accept or reject the testimony. The partial automation of scientific work raises new kinds of questions with respect to the attribution and distribution of epistemic responsibility. Who is epistemically responsible for the information provided by a scientific instrument? As computer technologies become more complex and behave increasingly autonomously, can or should humans be held responsible for the outputs of these technologies? Are machines capable of transmitting knowledge? What justifies beliefs based on machine outputs?

In this paper, I will be concerned with the last type of question. More specifically, the basic question I will be seeking to address is: How should belief-forming processes be individuated, for the purpose of epistemic assessment, in cases of epistemic reliance on instruments? In reliabilist epistemology—which I will be assuming throughout—answering the question of how to individuate belief-forming processes for the purpose of epistemic assessment is an important preliminary for understanding how the beliefs issuing from such processes can be justified. Sanford Goldberg (2012) has provided a clear answer to this question in the context of reliance on instruments: in such cases, belief-forming processes should not be extended to include the instrument. This is in contrast to reliance on human testimony where, he argues, the belief-forming process of the hearer should be extended to include the cognitive processing of the speaker. Goldberg’s reason for distinguishing between testimony-based and instrument-based belief is that the former, but not the latter, is a case of reliance on an epistemic subject. Human testifiers are subjects, and can therefore be expected to satisfy the same epistemic standards as the hearer. In contrast, instruments cannot be subjects (at least given the limitations of contemporary technology) and so cannot fully satisfy those standards. It is the continuity of normative expectations that justifies the extension in the human case.

In this paper, I will argue that the belief-forming process should include the instrument. According to the view I will defend, the proper unit of epistemic assessment, in cases of instrument-based belief, is not the isolated instrument but rather the instrument together with what I will call the 'instrument experts.' The latter category includes those whose expertise is based on their understanding of the instrument, for example the instrument-makers and professional operators. I will argue that the instrument together with the instrument experts jointly satisfy the same epistemic standards as human testifiers. Therefore, the belief-forming process of the recipient of instrument outputs should be extended to include the information-processing of the instrument/instrument expert pair. While this might seem to reduce to a case of reliance on human testifiers, I will argue that the epistemic opacity of much contemporary computer-based technology entails that epistemic assessment of beliefs based on such technology is irreducible to reliance on either the instrument experts or the instruments by themselves.

The paper is organized as follows. In the second section, I review Goldberg's argument for distinguishing between testimony- and instrument-based belief-forming processes, as well as some criticisms of his view raised by Kourken Michaelian. My approach to their disagreement will be to try to achieve a synthesis of their views. The topic of section 3 is the epistemic opacity of contemporary computer-based scientific technology. Drawing on the work of Paul Humphreys, I will argue that epistemic opacity makes the information-processing carried out by such technology irreplaceable by human information-processing in scientific work, and hence in the production of scientific testimony. In such cases, the expert testifier is dependent on the instrument for the propositional content of the testimony. In section 4, I discuss the "assurance view" of testimony and the role of assurance in expert testimony. In the fifth section, I argue that assurance is a feature of testimony that cannot be automated with current technology. Section 6 combines the results of the previous three sections to conclude that instruments should be included in belief-forming processes for the purpose of epistemic assessment. I offer concluding remarks in the final section.

2. Goldberg on Epistemic Extendedness, Testimony and Instrument-Based Belief

Goldberg (2010, 2012) defends what he calls the *Generic Epistemic Extended Mind Hypothesis*:

> GEEM: For at least some cases in which a subject *S* believes that *p*, a proper epistemic assessment of *S*'s belief requires an epistemic assessment of information processing that takes place in the subject's environment. (Goldberg 2012, p. 181)

Assuming a broadly reliabilist epistemology, Goldberg (2010) argues that GEEM is true of testimonial belief: the process of formation of the hearer's testimony-based belief includes not only the information-processing of the hearer, but also that of the speaker. This raises a natural question: If our reliance on other speakers in testimony cases is best understood as a case of interpersonally extended belief-forming processes, how should we understand cases in which we form beliefs through reliance on instruments and mechanisms, for example computers, clocks, gauges, thermometers, signal detectors of various sorts, and in general instruments on which humans rely for information about the world? Are these best understood as cases of what Goldberg calls "environmentally extended" belief-forming processes?

Goldberg rejects the extendedness model for beliefs formed through reliance on what he calls a "mere mechanism," such as a clock. The key difference between testimony-based and instrument-based belief, Goldberg argues, is that to rely in belief-formation on another speaker is to rely on an epistemic subject, which he describes as "a system which itself is susceptible to epistemic assessment in its own right" (Goldberg 2012, p. 182). In contrast, "mere" instruments and mechanisms are not properly regarded as epistemic subjects in their own right, and therefore they are not susceptible to normative epistemic assessment.

Goldberg's argument for applying GEEM to testimony is that it is implied by a plausible and more general principle:

> When a subject *S*'s belief that *p* is formed (or sustained) through a process π that takes as its input the output of a given stretch of cognitive processing π^*, then we should regard the belief-forming and -sustaining process relating to *S*'s belief that *p* as including both π and π^*. (Goldberg 2012, p. 183)

Memory turns out to be a special case of this principle: the assessment of a memorial belief should consider not only the reliability of the process by which the belief was sustained and retrieved but also that of the process by which the belief was initially produced. That is, memory is a *temporally* extended process. Likewise, Goldberg argues, testimony is another special case, in which the extension is *interpersonal* rather than temporal. For the reasons stated above, however, Goldberg does not think the principle holds for instrument-based belief.

Michaelian (2014) has criticized this restriction. According to Michaelian, Goldberg's argument consists essentially in the following two premises (Michaelian 2014, pp. 333-334):

1. Non-human components of distributed systems (cognitive systems consisting of humans and instruments) are not subject to normative epistemic assessment.
2. Extension depends on normative assessability.

Michaelian thinks both premises are problematic. With respect to (1), he argues that instruments are indeed subject to normative assessment:

> The difference between instruments and agents is that normative assessment of the latter is sensitive to a broader range of factors than is normative assessment of the former, not that the latter, but not the former, are subject to normative assessment at all. In the case of instruments, normative assessment tends to be sensitive primarily to relatively thin properties, such as reliability. In the case of agents, normative assessment tends to be sensitive to (in addition to these thin properties) various thicker properties, such as rationality and responsibility. But that does not make assessment of instruments any less normative, just as our assessment of agents does not become less normative when we take only their reliability into account. (Michaelian 2014, pp. 333-334)

With respect to (2), Michaelian claims that "employing normative assessability as the criterion for the extension of belief-forming processes is naturalistically unacceptable" (Michaelian 2014, p. 334). The reason he gives is that

> From a naturalistic point of view, the question whether a given belief-forming process is extended or not—the question of how to delimit the process—should turn on features of the process itself, not on whether we are prepared to assess it in normative terms. (Michaelian 2014, p. 334)

Summarizing Michaelian's objections, we can say that with respect to (1), the problem is that instruments are indeed subject to normative epistemic assessment, and that with respect to (2), the problem is that normative features are not part of the belief-forming process, and therefore are incompatible with naturalism.

I disagree with both objections. The first, in my view, is based on a misunderstanding of Goldberg's argument. That "normative assessment of [agents] is sensitive to a broader range of factors than is normative assessment of [instruments]" is key for Goldberg. Though he does not state the point explicitly, Goldberg seems to think that three norms are especially relevant for epistemic assessment: reliability, rationality and responsibility, what I will call the "three Rs." His point concerning instruments is that "the only relevant assessment is whether the information processing done by the instrument was reliable, not whether it was irrational or irresponsible" (Goldberg 2012, p. 188). True, as Michaelian observes, early in his (2012) Goldberg does write that instruments "are not susceptible to normative epistemic assessment" (p. 182, cited above), but later in the article he qualifies this restriction by excluding instruments only from the "*full* normative assessment of (the doxastic states of) epistemic subjects" (p. 188, emphasis added). He then states what this full normative assessment consists of:

> Let α be any causal sequence that eventuates in an "output" o, where o can be regarded as the output of an information-processing sequence, and where this output itself is such that some subject might "rely" on this output in belief-formation … Then we can say: α's information processing is relevant to assessments of the doxastic justification of beliefs formed through reliance on α's output if, but *only if*, (i) α's information processing can be assessed for reliability *and* (ii) α's operations are properly assessed for rationality and responsibility.

Goldberg claims that cognitive-psychological processes such as those involved in human thinking count as cases in which (ii) holds, whereas information processing performed by ordinary instruments does not. As a shorthand, I will call information processing that only satisfies (i) "1R-assessment" and information processing that satisfies both (i) and (ii) "3R-assessment."

Michaelian seems to think that the difference between 1R- and 3R-assessment is of minor importance, whereas Goldberg thinks the contrary. Goldberg's reason for thinking the difference matters (though he does not state the point in the general terms that follow) seems to be that ignoring or downplaying the difference between the two kinds of assessment has the effect of making inferences, the proper justification of which would require significant background assumptions, appear justified even without those assumptions. He gives the example of a subject *S* who has a disposition such that whenever she sees a tree stump with *n* rings, she forms the belief that the tree is *n* years old. This even though she knows nothing about the covariation between the number of rings on a tree and the tree's age, or about the natural laws responsible for the covariation. The process by which she comes to form beliefs about trees is reliable, of course, but nevertheless her uncritical reliance on it is irresponsible because she has no evidence for the inference from number of tree rings to tree age.

Note the nature of the process: it consists of inferring a belief about a state of affairs (the tree's age) not directly observed from a state of affairs (the rings) directly observed. Even though reliable, the process seems unjustified without the appropriate background knowledge. As Goldberg proceeds to argue, such unjustified inferences closely resemble perceptual belief formed under conditions in which the perceiver has no independent reasons for regarding perception as reliable. But intuitively, such perceptual beliefs are responsibly formed (people and perhaps some animals form them all the time), which suggests that the inferences are responsibly formed as well. But this result conflicts with another intuition, which is that inferential belief requires justification that perceptual belief does not.

Though more could be said to support taking the distinction between 1R- and 3R-assessment seriously, I will assume here that enough support has been provided to make a version of premise (1) above, suitably revised in light of the preceding discussion, seem more reasonable:

3. Non-human components of distributed systems (cognitive systems consisting of humans and instruments) are not subject to 3R-normative epistemic assessment.

(3) merely qualifies (1) by inserting "3R" to make more precise the respect in which the non-human components fail to satisfy the standards of epistemic assessment for subjects.

I will now briefly suggest a line of response to Michaelian's second objection. In the passage quoted above, he implies that normative features are not part of the belief-forming process, and therefore are incompatible with naturalism. The point, I take it, is that an epistemic subject cannot be a feature of the process, presumably because such an entity is incompatible with a thorough-going naturalism. Though it is beyond the scope of this paper to assess the point just stated, it seems reasonable to suggest that although one would like one's epistemology to be compatible with naturalism, one would also like one's naturalism to be compatible with the modes of functioning of Western cultures. In these cultures, individual humans are assumed to be loci of agency, subject to attributions of rationality and responsibility. The social, legal and moral systems of Western societies depend on this assumption.[17] Thus for any version of naturalism that is sensitive to the social contexts in which belief-forming processes occur, I do not find it "methodologically unacceptable," as Michaelian does (p. 334), to make normative assessability the criterion for the extension of the

[17] Giere (2012) defends a model of free will, which is supposed to be compatible with scientific naturalism, along similar lines.

belief-forming process. Though I agree with him that more needs to be said for a full defense of 3R-assessment as the criterion for extension, I do not believe that the objections he poses *prima facie* exclude making use of Goldberg's criterion, and so I will proceed to make use of it for the remainder of the paper.

The reason I have discussed the difference of views between Goldberg and Michaelian in some detail is that I think both are partially right and partially wrong about what elements should be included in belief-forming processes, for the purpose of epistemic assessment. One of my goals in the following sections is to try to achieve a sort of synthesis. My view is as follows:

4. Against Michaelian, and in agreement with Goldberg, there are good reasons for including only 3R-assessable processes in belief-formation.
5. Against Goldberg, and in agreement with Michaelian, instruments should be included in belief-forming processes for purposes of epistemic assessment.
6. Against Michaelian, acceptance of (5) is compatible with (4).

On Michaelian's view, claim (6) is false, since instruments are not 3R-assessable on that view (*vide supra*). On the other hand, despite my broad agreement with Goldberg about the desirability of the "3 Rs" for individuating belief-forming processes, I do not think these norms entail the exclusion of instruments from those processes. The reason has to do with the irreplaceability of instruments and experts in the epistemic assessment of certain kinds of contemporary computer technology, a situation that will be the general topic of the next three sections. I will assume here that enough has been said in this section and by Goldberg himself for (4) to be considered plausible. The next two sections will be concerned with providing grounds for (5). The defense of (6) will come in the penultimate section of the paper.

3. Appeals to expert opinion

On the face of it, instrument-based belief and testimony-based belief may seem very different, due to the difference between the sources of belief. Yet there is a growing trend in various social practices to use the two sources for the same purpose, namely, as the basis for an appeal to expert opinion. In the introduction, I mentioned a real case in which the source of a type of claim about the provenance of DNA crime samples, which source would formerly have been the human analyst, was now a fully automated process. The argument of this and the following section is that, in the production of contemporary scientific testimony, human and machine are becoming mutually irreplaceable. This argument can be summarized in the slogan, "assurance can't be automated, computation can't be manualized." In this section, I provide grounds for the plausibility of the slogan's second claim. The significance of this claim for scientific testimony is that the impossibility of "manualizing" the information-processing that is required to testify that *p* entails that if the expert wants to testify that *p*, the expert is dependent on the information-processing in the machine for the propositional content of her testimony.

It will be helpful to keep in mind the context in which appeals to scientific opinion, or the outputs of scientific instruments, are made. Within science, instruments are generally applied by, and for the benefit of, scientists who are experts themselves in the field of application of the instrument. For example, a chemist who uses a spectrometer to study molecular structure is presumably an expert in the field of application of the spectrometer, molecular structure. Scientists may even, and frequently have been, experts in the subject domain constituted by the instrument itself: for example, some chemists have also developed the spectrometers they used.

In appeals to scientific opinion, however, an essential feature of the context of use is that a scientific claim is intended to persuade non-experts, for example the members of a jury. An important question, therefore, is what level of understanding the non-experts should have of the information processing used to arrive at the claim, in order for their evaluation of it to be rational. In the case of instruments, at one extreme the instrument may be no more than a black-box to them, that is, may be understood by them as merely an input-output device. The inner workings of the instrument and the various ways it interacts with the sample, environment and human operators are unknown to them, or at most they have only a drastically simplified notion of these. At the other extreme, the system may be a “white-box.” In such cases, a representation of the inner workings of the instrument is available. Typically, the internal process of the machine may be represented as a collection of modules rather than as a single input-output unit. Each module is characterized by one or more state parameters, laws of temporal evolution, and laws of interaction with other modules. The collection of modules and laws constitutes a more detailed model of how the system works than a black-box model.[18]

So in the case of appeals to instrument outputs, the level of understanding that the targets of the appeal should have is an important consideration. In cases where a thorough, “white-box” understanding of the system is in order, such an understanding may be beyond the ability of the layperson to achieve. This predicament is especially likely with sophisticated technology, such as the DNA genotyping software mentioned in the introduction. In such situations, it will most likely be necessary to delegate an expert to come to an understanding of the system, and communicate enough of his understanding of how it works to the non-experts for the latter to evaluate the appeal in a rational way. An example of such delegation is again provided by the genotyping software, where the defense and prosecution delegated rival experts to scrutinize the software.

Whether the non-experts come to an understanding of the instrument themselves, or with the help of an expert, it remains an important question how open to critical scrutiny the instrument has to be in order to evaluate the appeal to its output. It is worth noting that the instrument need not be a white-box in order for its reliability to be assessed. For example, a well-known method for testing black-boxes is calibration, where the instrument under test

[18] See Tal (2012) for a discussion of the differences between black-box and white-box models.

reproduces known results. In some cases, such methods may be sufficient to warrant the appeal to the instrument's output, but in others it may be necessary to "open the black-box."

Similar considerations apply to expert opinion. Human experts, when forming their opinions, often engage in complicated reasoning, drawing on esoteric subjects, that laypersons cannot be expected to understand. The main difference is that an expert can answer questions concerning his or her expertise and how he or she reasoned to the opinion. In contrast, instruments cannot answer questions about their own inner workings or their past results (unless they have been specifically programmed to do so). On the other hand, instruments are designed and constructed by humans who are available, at least in principle, to answer such questions in their stead. The availability of human experts raises the question whether instrument output is reducible to expert opinion. Surely the instrument experts can answer questions about how the system works, why the output should be trusted, how its past results have been validated, etc.? In such cases, the instrument's output can be supplemented by the expert's account of how the instrument works. Such cases may reduce to the ordinary appeal to expert opinion, in that what is ultimately happening is that the expert is giving an opinion as to why a result produced by the machine should be accepted.

A word on my use of the term 'reduction.' The term has been extensively used in various philosophical contexts, and so I would like to be clear about what I mean by it. The idea is that, at least in principle, the information-processing of the instrument could be replaced by that of the human scientist. I include only information-processing, because clearly there are physical functions that humans could never replace. For example, instruments based on signal detection have detectors that far outstrip the detection limits of the human senses. By focusing on information-processing, my target here is a tendency in traditional philosophy of science to completely abstract from practical constraints. For example, Manfred Stöckler suggests that "in principle, there is nothing in a simulation that could not be worked out without computers"

(2000, p. 368).[19] As Paul Humphreys (2009) points out, however, "almost all contemporary simulations require abilities that go far beyond what is possible by the unaided intellect" (p. 623). A prime example are Monte Carlo simulations, such as that employed by TrueAllele (Perlin 2013, p. 2).

These are cases where the human expert is unable to provide an account of how the system works, or of what Humphreys (2004, 2009) calls 'epistemic opacity.' Humphreys (2009) identifies two kinds of epistemic opacity. In epistemic opacity *simpliciter*, a process is epistemically opaque relative to a cognitive agent X at time *t* just in case X does not know at *t* all of the epistemically relevant elements of the process. A stronger sense of epistemic opacity is what Humphreys calls 'essential epistemic opacity:' "A process is essentially epistemically opaque to X if and only if it is impossible, given the nature of X, for X to know all of the epistemically relevant elements of the process" (Humphreys 2009, p. 618). Such a process is no longer fully decomposable into modules that can be understood individually and in combination, unlike the white-box models mentioned earlier. Humphreys has focused on cases of computer simulations where the computations involved are "so fast and so complex that no human or group of humans can in practice reproduce or understand the processes." For example, some computer-assisted mathematical proofs are so complicated that no human can comprehend them.[20] Another type of case involves computer programs where the source code is opaque for some reason. Perhaps there are social barriers to accessing the code, for example intellectual property rights in the case of commercial software. Thus social context may be an important determinant in the epistemic opacity of a process. Or the code may be so massive that it becomes difficult, if not impossible, for an individual human to comprehend it.

[19] Cited in Humphreys (2009), p. 623.

[20] The rise of computer-assisted proofs has generated debates within mathematics and philosophy of mathematics over the desirability of "surveyable" proofs. See Mackenzie (2001) for a book-length treatment of the rise of "mechanized" proofs.

The existence of epistemically opaque systems raises a challenge for the view, mentioned above, that the appeal to instrument output is reducible (in the sense specified above) to the appeal to expert opinion. If a process is epistemically opaque in the sense that "no human or group of humans can in practice reproduce or understand" it, then even an expert on the system will not be able to reproduce or understand the process by which the system arrived at its result. Furthermore, given the number of degrees of freedom (up to 10^6 and more) and the number of computational steps (sometimes billions) that are handled by computer simulations (Humphreys 2004), it would not be feasible to replace the process by a different method that a human could carry out. So in practice, if we want the results such systems afford, we have no choice but to use epistemically opaque systems.

Not only does epistemic opacity challenge the reducibility view, but it introduces a *prima facie* distinction between using the testimony of human experts and using the outputs of instruments. If it is granted that a human expert is in conscious control of his or her reasoning process, then the recipient of expert testimony is entitled to expect that the expert could provide an account of the steps in his or her reasoning that led the expert to the conclusion. In contrast, an account of every element of the computational processes that produce the outputs of an essentially epistemically opaque system is inaccessible to humans, even the instrument experts. One consequence of this inaccessibility is that it would appear to block a major avenue for evaluating appeals to expertise, namely the asking of questions in order to reconstruct the expert's reasoning.

4. The role of assurance in (scientific) testimony

In this section I argue that scientific testimony has a dual character. Consider an expert who is asked to support a position on a matter of dispute (a trial, public policy, etc.). On the one hand, her testimony is the utterance of propositions based on her own information processing and whatever facts serve as the initial inputs to it. The fact that a proposition *p* is uttered, by an expert in the domain of *p*, is evidence for *p*'s truth. On the other hand, it is an explicit assumption of responsibility for the truth of the proposition, in which the expert presents herself as a guarantor of its truth. So the expert also *assures* the audience that *p* is true. Thus the expert's utterance provides both evidence and assurance that *p* is true. Let us examine each aspect in turn.

As the utterance of a proposition based on evidence, the testimony can be epistemically evaluated in light of the evidence for it, which includes not only the expert's own information processing and input data, but also such corroborating evidence as whether it is consistent with the opinions of other experts and whether the technical reasoning of the expert actually supports the position she was brought in to support. One characteristic of the use of scientific testimony in extra-scientific contexts is that technical claims are often used to support non-technical claims, like verdicts of innocence or guilt, or medical claims concerning lifestyle ("smoking is harmful to one's health"). So for purposes of epistemic assessment it is sometimes a legitimate question what exactly the expert asserted that implies the final claim, support for which is sought by appealing to the expert's testimony. A general problem that the question reveals is that experts can be misunderstood, as a result of distortion, misquotation, or misinterpretation. Another problem is that expert opinions tend to be guarded, fairly technical claims that are difficult to accurately represent in layperson's terms. For these reasons, a respondent to scientific testimony is entitled to ask whether what the expert actually asserted entails the conclusion that is being supported by appealing to the expert.[21]

Another avenue for epistemic assessment is to inquire into the process by which the expert reasoned from the data to her opinion. Experts are fallible, of course, and so knowing how the expert produced her opinion allows the respondent to check for flaws in the information-

[21] See Walton (1997) for an analysis of the issues involved in evaluating appeals to expert opinion.

processing. Moreover, the processing will sometimes involve controversial moves or the exclusion of alternative methods that may affect the content and reliability of the assertion. Questions about the information-processing, if answered, provide the respondent with justifications for the choices the expert made in coming up with her opinion.

It should be kept in mind, however, that we are assuming throughout this discussion that the expert is testifying to non-experts. So it is unlikely that all of the participants besides the expert will be able to fully understand the information-processing by which he arrived at his opinion, or the technical claims from which the opinion is inferred (if the two are distinct). This situation of incomplete comprehension contrasts with the situation of testimony given by the scientist to his colleagues where it is reasonable to assume that, at least in ordinary cases, all of the information-processing can be made fully explicit and can be fully understood by all the participants. Furthermore, it is also possible that even the expert cannot make the process by which he arrived at the opinion fully explicit, given the possible role of intuition and implicit assumptions in the exercise of expertise.[22] In general, the other participants cannot be expected to adequately evaluate the expert's opinion based purely on their understanding of his information-processing. So some assurance is needed for the testimony's truth that goes beyond the evidence for it. This assurance is provided by the fact that the expert *asserts* the proposition, and that he asserts it *as an expert*. By asserting the proposition, he guarantees its truth, and the guarantee is meaningful because the speaker is an expert in the domain of the proposition.

The remarks in the preceding paragraph are based on the "assurance view" of testimony, according to which the epistemic import of testimony cannot be reduced to that of being evidence for the truth of the propositions constituting the testimony. According to what Moran

[22] Walton (1997, p. 110) calls this the "inaccessibility thesis."

(2005) calls the Evidential View, the epistemic import of testimony consists in the fact that it is evidence for the truth of a proposition, much like a photograph or a "natural sign" like a footprint might be evidence for the truth of a proposition. Moran calls such a view "reductionist" and argues that it misunderstands the nature of testimony. In contrast to this view, Moran claims that in testifying,

> The speaker is asking that a certain authority of his be acknowledged—the authority to invest his utterance with a particular epistemic import—and this investment occurs by his explicit assumption of responsibility for his utterance's being a reason for belief. (Moran 2005, p. 18)

If the speaker's authority is accepted, then the utterance is endowed with epistemic import that is independent of the evidential relation between it and the truth of the proposition uttered. On this view, testimony is similar to an agreement or contract, in which one party takes responsibility for the truth of a proposition and where the other acquires the right to complain if that responsibility is not satisfied. The legitimacy of the speaker's authority is, of course, dependent on whether certain background conditions are met, such as whether she has the relevant knowledge, abilities, trustworthiness and reliability.

Importantly, if the audience accepts the speaker's authority, this authority gives the audience a reason to believe *the speaker*, and not just her utterance. Indeed, the rationality of believing the utterance is now derived from that of believing the speaker. The priority of belief in the speaker results from the fact that she has publicly assumed responsibility for the truth of her utterances by engaging in the act of assertion.

Moran (2005) does not consider the phenomenon of scientific testimony, but what he argues for testimony in general would seem to follow *a fortiori* in the case of experts. Moran's paper is motivated by the desire to understand how it is possible for us to depend rationally on what others tell us, given all the ways in which human testimony is fallible (being subject to deceit, error, carelessness and other flaws.) In addition to these general problems with testimony, scientific testimony has specific features that add further pitfalls in appeals to expert opinion. Perhaps the most important of these is the lack of understanding possessed by non-experts. The more complicated or esoteric the expertise involved, the more important will be the expert's assurance in evaluating the conclusion being argued for. In addition, there are also the aforementioned problems that the technical claims of the expert do not necessarily entail the conclusion being argued for by means of them and that it may be impossible to make the evidence fully explicit. So if, in the case of ordinary testimony, the speaker's assurance, together with his satisfaction of the background conditions, can be a good reason to believe what he asserts, then these barriers to evaluating scientific testimony may make acceptance of the latter on the basis of the expert's assurance the only option (besides rejection).

Since the expert's assurance is only warranted if she meets the background conditions for being an expert in the domain of the conclusion being argued for, the targets of the assurance are entitled to check whether these conditions have been met. This entitlement underlies the practice of asking question concerning the expert herself (as opposed to her assertions), for example during cross-examination in court. Questions concerning the credibility of the scientist as an expert source, her personal reliability, or the relevance of her expertise to the matter at hand can all be interpreted as aimed at checking whether the background conditions for warranted assurance have been met.[23]

5. Assurance can't be automated

In general, expert testimony is never fully automated, as the production and use of the machine requires human experts that produce, operate and make it their business to understand it.[24] But incomplete automation is also to be expected in light of the dual character of expert testimony

[23] See Walton (1997), Ch. 7 for a discussion of the appropriate "critical questions" to ask when evaluating expert testimony.

[24] An expert need not be always present with the machine, of course. Non-experts can learn to use scientific instruments for their purposes. I refer here to the *community* of instrument experts, members of which may or may not be present for any particular instance of machine use. I will have more to say about the nature of the instrument experts in the final section of the paper.

discussed in the previous section. As evidence, the expert's testimony is the result of a process of reasoning from input information to the conclusion (or to a technical claim from which the conclusion is inferred). As assurance, the expert's testimony represents the assumption of responsibility for the truth of the conclusion (or of the aforesaid technical claim). I submit that the former, but not the latter, can be automated, at least given the current state of technology. The reason is that to treat the speaker as responsible for the truth of her utterances is to treat her as an epistemic subject who can be held answerable to the standards of proper testimony. This involves, among other things, treating her as the proper subject of blame if the testimony turns out to be inaccurate. But machines are not the sort of thing we hold epistemically accountable, at least given the current state of technology. Even in cases where our most sophisticated computers yield information that turns out to be false, it is the humans who produce and use it that will be blamed, not the computer itself.

It is worth noting that Goldberg also thinks that assurance constitutes an essential difference between human testimony and instrument output:

> It is worth underscoring the difference in this regard between a speaker's assertion that *p*, and a state of an instrument (such as the state of a clock's face) that represents that *p*. There can be little doubt that if a state of a mechanism is to count as representing that *p*, there must be some sort of relation that holds between the mechanism's being in that state and the obtaining of the state of affairs in which *p* … In light of obtaining of this relation, we might say that the mechanism's being in that state is (properly regarded as) *evidence* that *p*. And if the instrument was designed with this purpose—so that it is being [*sic*] in a certain state could be recognized as indicating that *p*, or being evidence for *p*—we might even describe the mechanism's being in that state as its *representing* that *p*. But even so, the speech act of testifying contrasts with this. As many authors (citing Grice 1957 as inspiration) have noted, one who testifies that *p* is not intending that her speech act should be taken as evidence that *p*, or as merely a signal indicating that *p*; rather, she is herself vouching for the truth of *p*. (Goldberg 2012, p. 192)

In an endnote, Goldberg states that he borrows the language of vouching from Moran (2005). In the scientific scenario we have been envisaging, the instrument does indeed "represent that *p*." In the DNA case mentioned in the introduction, for example, the computer represented that the source of the DNA evidence was 5.7 billion times more likely to be the suspect than coincidence. The representational role of the instrument does not exclude, however, the existence of an agent who can vouch for the truth of *p*. There *is* such an agent—the instrument expert.[25]

My suggestion is that in these hybrid uses of machines and human experts to provide scientific testimony, the epistemic functions of assurance and evidence are no longer carried out by the same individual. In effect, what has happened is a distribution of functions between human and machine. This distribution is represented in the following diagram:

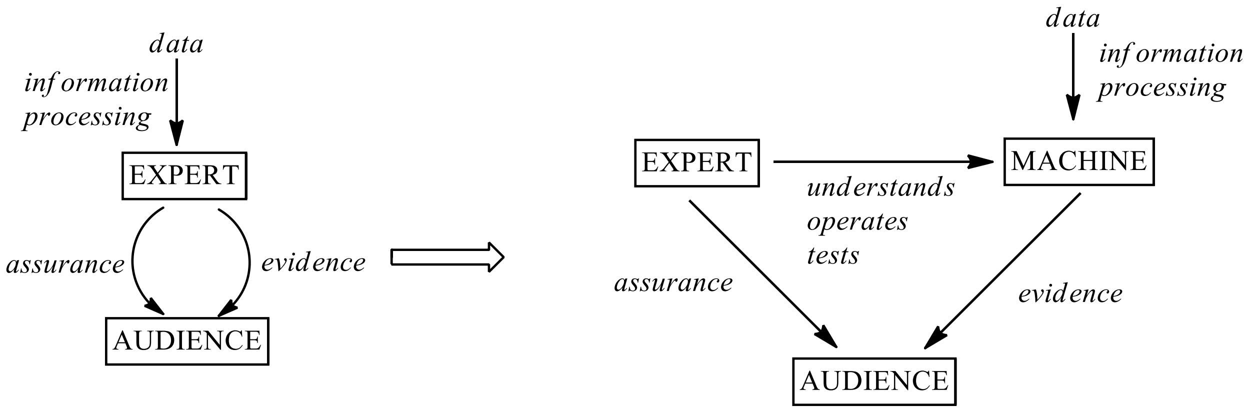

Figure 1. The distribution of functions between the human expert and the machine.

At this point, one might ask whether, on the assurance view, human testimony that *p* is *ever* evidence that *p*. Though the wording is somewhat ambiguous, the passage from Goldberg quoted above suggests that it may not be, that the epistemic import of assertion is exhausted by vouching. Moran (2005) is also somewhat ambiguous on this point, though for the most part he seems more concerned with attacking the opposite view, that the epistemic import of testimony reduces to evidence, and at certain points seems to brush with the idea that testimony might serve as both assurance and evidence (e.g., p. 23). As an admittedly tentative answer, my suggestion is to focus on the pragmatics of appeals to expert opinion. Walton (1997, Ch. 7) identifies six critical questions that an audience is entitled to ask, and that audiences do in

[25] In the case reported in Ward (2016), for example, the reliability of the genotyping program's outputs was vouched for by the software developer as well as by external expert users of DNA genotyping software.

fact ask in social contexts like trials, when assessing expert testimony. Of these, three concern the expert herself, and three concern her assertion. My suggested explanation for this division is precisely that expert testimony has the dual character I have attributed to it. On the one hand, it is evidence of a state of affairs in which *p*. On the other, it is a guarantee of the truth of *p*. The former is corroborated by independent evidence for *p*, whereas the latter is corroborated by evidence for the credibility, reliability, knowledgeability, and so forth of the expert.

In the relationship between expert, machine and audience indicated in Figure 1, a key role is played by the expert's understanding of the machine, for it is this understanding that links the evidence and the assurance. If the machine is epistemically opaque to the expert, then the value of the latter's assurance may seem doubtful. This source of doubt raises the question of what exactly the assurance is based on, and whether it can be based on a less than complete understanding of the computational process. One might also ask whether any assurance is necessary. Can the machine's output stand on its own? But allowing this would put laypersons in the position of having to accept its claims at face value, much like the deliverances of an oracle or clairvoyant. But acceptance of machine output at face value borders on irrational, because acceptance of such outputs is supposed to be supported with evidence, for example that the instrument is calibrated, that its design is based on a reliable theory of the instrument, that noise has been appropriately dealt with, etc.

Or perhaps something like assurance can be provided by the machine itself. But the provision of assurance by the machine would require the existence of ways that might be used to establish that a machine is to be regarded as an epistemic subject in its own right. The main question here is what sort of behavior the machine would have to display such that it would be properly assessable not just in terms of reliability, but also rationality and responsibility. As far as I know, computer science and robotics have yet to produce a machine that convincingly exhibits such behavior.

But perhaps one could argue that requiring the machine to be an epistemic subject is too demanding. Such a view would be plausible were the epistemic benefit of assurance recoverable without a subject. One possibility, for example, is that the epistemic benefit is supervision and control (Green 2006, Ch. 2). By assuming responsibility for the truth of her conclusion, the expert is effectively assuring the hearer that her information processing was properly supervised and controlled. By "properly supervised and controlled," I mean that the expert took care that the information-processing satisfied whatever rational requirements are needed for the process to be truth-conducive. On this view, the dual character of testimony would simply reflect the fact that epistemically responsible information-processing in fact involves two processes. On the one hand, there is a "first-order" process of reasoning from data to conclusion. On the other, there is a "second-order" process of supervising and controlling the first-order process.

In the case of humans, the second-order process requires self-consciousness, in other words supervision by a conscious mind. But if the expert's reasoning is replaced by a computational process, there may appear to be no reason why the process of supervision of the computation must be conscious. The mere fact that a process of supervision is conscious does not, by itself, create or impair any epistemic benefit from the point of view of the conclusion issuing from the information-processing (Green 2006, p. 37).

If consciousness is indeed inessential to the process of supervision, then it may seem as if there is no obstacle, in principle, to full automation of expert testimony. The problem, in my view, is that the epistemic benefit of assurance is not limited to a guarantee of supervision and control. As noted above, assurance is supposed to entitle the hearer to blame the testifier when the testimony is found to be deficient. It is also supposed to entitle the hearer to "pass the buck" to the speaker when the hearer transmits the testimony to a third party.[26] The reduction of the epistemic benefit of assurance to supervision and control cannot account for these entitlements, because a mere mechanism of supervision and control cannot be the subject of blame or passing the buck. Thus if automation is to proceed without loss of these entitlements to the hearer, it must in some way recover the epistemic subject. As mentioned above, I do not believe this is possible with current technology.

6. Instruments should be included in belief-forming processes for purposes of epistemic assessment

In summary, the proper way to understand instruments is not as "mere" mechanisms, but as the result of a distribution and partial "externalization" of functions ordinarily carried out by epistemic subjects. Humans, for example, detect causal inputs from the world through the senses and process the information by the use of their rational faculties. Furthermore, they carry out second-order processes of supervision and control to ensure the fulfillment of the rational requirements for the information-processing to be truth-conducive. They may also answer to attributions of blame and buck-passing. Under certain conditions, including important technological ones, humans can choose to delegate certain of these processes to machines. With sufficient technological development, machines may be developed with capabilities that far outstrip native human abilities, including the ability to understand the information-processing carried out by the machine. At this point, the machines become irreplaceable in practice, for they carry out information-processing tasks that humans cannot.

Given the extremely powerful evidence such machines give humans access to, it becomes highly desirable to use this technology in place of human testifiers. To return to the example of the DNA genotyping program, whereas the program can analyze complicated DNA samples

[26] On the phenomena of blame and buck-passing and their connection with testimony, see Goldberg (2016), Ch. 3.

and afford likelihoods on the order of one in billions (at least), the methods used by human analysts only afford likelihoods on the order of one in thousands (Perlin 2013, p. 1). On the other hand, given the current state of technology, the epistemic functions of assurance and evidence have to be divided between humans and machines for the foreseeable future. On the view defended in the previous section, assurance remains necessary for the rational and responsible production of the instrument-based outputs. Furthermore, from the point of view of the *reception* of these outputs, the existence of human experts who provide assurance makes deference to the outputs, by laypersons, rational. So humans are irreplaceable as well, at least according to the view of testimony adopted here.

How does all of this help us with the question we started with, how belief-forming processes should be individuated in cases of epistemic reliance on instruments? Recall the general principle from which GEEM is drawn:

> when a subject *S*'s belief that *p* is formed (or sustained) through a process π that takes as its input the output of a given stretch of cognitive processing π^*, then we should regard the belief-forming and -sustaining process relating to *S*'s belief that *p* as including both π and π^*. (Goldberg 2012)

I have been arguing that the process of machine-based scientific testimony is in fact a dual process, in which the functions of cognitive processing carried out by humans in ordinary cases of testimony are now distributed between the scientist and the machine. I have further argued that both the scientist and the machine are necessary for the output of the dual process to count as testimony. Importantly for the overall argument of this paper, the scientist and the machine jointly satisfy Goldberg's three norms of epistemic assessment: reliability, rationality and responsibility. Considered together, then, the instrument/instrument expert pair is 3R-assessable. From this perspective, beliefs based on the output of such pairs are no different than beliefs based on memory or ordinary testimony. Since these were special cases of the general principle, I conclude that machine-based scientific testimony is also a special case. It follows that GEEM applies to such testimony, or in other words, that the extendedness model applies to this kind of reliance on instruments. In terms of the discussion in section 2, then, there is no incompatibility between claim (4)—the requirement of 3R-assessability—and claim (5), the inclusion of instruments in belief-forming processes.

7. Concluding remarks

The argument of this paper is essentially for the conditional claim that *if* extended reliabilism (GEEM) is correct, and *if* the assurance view or a sufficiently similar view of testimony is correct, then belief-forming processes should include machine-based processes of the sort described above. Both antecedents have been subject to criticism, so whether the consequent is necessarily true will depend on whether that criticism has been successful in undermining either extended reliabilism or the assurance view. I will not try to assess the existing criticism of those views in this paper, but merely flag this task for future work.

Of course, my application of the assurance view to machine-based scientific testimony is itself hardly immune from criticism, though I hope to have provided some reasons for thinking the idea non-crazy. In lieu of a conclusion, I would like to make a few additional comments on the nature of the 'instrument experts' I have been referring to throughout this paper, in order to indicate a way in which the views defended here might be applicable in real-world cases.

The following dilemma may be extracted from the phenomenon of epistemically opaque technology. Current automation technology gives us access to extremely powerful evidence not otherwise available. But the increasing complexity and speed of computational processes render their operation increasingly opaque to human experts, threatening the ability of the latter to provide assurance for the systems' outputs. But such assurance is, on the view defended here, necessary both for the validity of the appeal to the outputs and the preservation of epistemic responsibility. As mentioned above, it is what makes deference to the outputs of the technologies rational. So we appear to be confronted with a dilemma: either we reject such technology, and lose the powerful evidence it provides, or we accept it, and run the risk of creating a society in which deference to experts loses its rational basis.[27]

I have made two basic assumptions in my treatment of assurance in this paper. First, I have assumed that assurance is provided by individual experts rather than groups. Second, I have assumed that assurance reflects an individual process of supervision and control by an epistemically responsible agent. My suggestion for avoiding the dilemma is to relax both assumptions.

This suggestion is based on my more general view that the concept of assurance needs to be adjusted to reflect the nature of the technology employed by the expert. The sophisticated machines of contemporary science result from the combination of advances from a broad range of scientific and technological fields. The contemporary scientific instrument generally results from the work, over long periods of time, of many people, whose combined expertise is concentrated in the instrument. Furthermore, contemporary scientific technology relies heavily on complex and extremely rapid computational processes that make the technology epistemically opaque.

The nature of contemporary scientific technology, as just sketched, has two consequences for the concept of assurance. First, the individual scientist may be increasingly ill-suited to be the guarantor of the instrument's reliability. A better guarantor may be the scientific community whose combined expertise has enabled the production, testing and understanding of the instrument. Though in some cases no group of humans, let alone an individual human, can understand the details of the instrument's processes, it seems likely that in such cases more progress can be made towards understanding through a community-based approach than an individual one. Second, the assurance that the community can offer at any given time is the

[27] On how deference to experts can be rational, see Hardwig (1985).

result of a spatially and temporally extended process in which the instrument has been studied and checked by members of the community. It can no longer be up to the individual scientist to take care that the processes resulting in output are such as to satisfy the rational requirements for the process to be truth-conducive. The need for collective supervision, control and responsibility results both from the complexity of the instruments, but also the fact that certain flaws only reveal themselves over time. This latter feature is especially characteristic of computationally-based processes. For example, bugs in computer programs sometimes only manifest themselves after long periods of time. Therefore, calibration studies performed on the programs during narrow periods of time may fail to detect all the bugs. The collective study and use of computer software may accelerate the process of bug detection.

So one question for further investigation is whether collective scrutiny is indeed necessary to provide assurance, or whether other methods may be capable of doing so? This question is interesting at least in part because if the first disjunct were true, this would amount to the demand that the technology be made as transparent as possible. Transparency, however, may conflict with other social practices, like trade secrecy and intellectual property rights. So one avenue for further research would be to investigate to what extent the collective character of the assurance required for contemporary technology, if it is required, can be reconciled with social practices that are resistant to "collectivization." Another question is how exactly to delimit the scope of the category 'instrument experts.' Does it involve only scientists operating machines in actual cases, or does it involve the instrument-makers, designers and theorists of the instrument, and others who may not be present in any actual applications of the technology? What level of understanding of the technology must a person have in order to qualify as an expert? If assurance requires a temporally extended and indefinite process of de-bugging, at what point can assurance be given by the expert that the machine's output is reliable?

8. References

Giere, R. N. (2012). Scientific cognition: human centered but not human bound.

Goldberg, S. C. (2016). *Assertion: On the Philosophical Significance of Assertoric Speech*. Oxford: Oxford University Press.

—(2012). Epistemic extendedness, testimony, and the epistemology of instrument-based belief. *Philosophical Explorations*, 15(2): 181-197.

—(2010). *Relying on Others: An Essay in Epistemology*. Oxford: Oxford University Press.

Green, C. R. (2006). The Epistemic parity of testimony, memory, and perception. Ph.D. dissertation, University of Notre Dame.

Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82(7): 335-349.

Humphreys, P. (2009). The Philosophical novelty of computer simulation methods. *Synthese*, 169: 615-626.

—(2004). *Extending ourselves: Computational science, empiricism, and scientific method*. New York: Oxford University Press.

—(1994). Numerical experimentation. In P. Humphreys (Ed.) *Patrick Suppes: Scientific Philosopher, Volume 2. Philosophy of Physics, Theory Structure and Measurement Theory*, pp. 103–118. Dordrecht: Kluwer Academic Publishers.

Mackenzie, D. A. (2001). *Mechanizing Proof.* Cambridge, MA: The MIT Press.

Michaelian, K. (2014). JFGI: From distributed cognition to distributed reliabilism. *Philosophical Issues*, 24: 314-346.

Moran, R. (2005). Getting told and being believed. *Philosophers' Imprint*, 5(5): 1-29.

Perlin, M. W. (2013). Cybergenetics TrueAllele technology enables objective analysis of previously unusable DNA evidence. www.mathworks.com. Accessed December 14, 2016.

Pinto, M. F. (2016). Economics imperialism in social epistemology: a critical assessment. *Philosophy of the Social Sciences*, 1-30.

Stöckler, M. (2000). On modeling and simulations as instruments for the study of complex systems. In M. Carrier, G. Massey, & L. Ruetsche (Eds.), *Science at Century's End: Philosophical Questions on the Progress and Limits of Science*, pp. 355–373. Pittsburgh: University of Pittsburgh Press.

Tal, E. (2012). *The Epistemology of Measurement: A Model-Based Account*. Dissertation, University of Toronto, Graduate Department of Philosophy.

Walton, D. (1997). *Appeal to Expert Opinion.* University Park: The Pennsylvanis State University Press.

Ward, P. R. (2016). Legal question: how do you cross-examine a computer? *Pittsburgh Post-Gazette*, August 29th.