Paul Humphreys

Network Epistemology[1]

ABSTRACT

A comparison is made between some epistemological issues arising in computer networks and standard features of social epistemology. A definition of knowledge for computational devices is provided and the topics of nonconceptual content and testimony are discussed.

In this paper I shall address some basic issues in the epistemology of networked computers. More particularly, I shall highlight some differences between standard social epistemology and the epistemology that is appropriate for computational networks, and discuss how such networks can serve as a source of testimony. In addition, I shall explore the role content plays in the interface between networks and humans, whether the content be conceptual, non-conceptual, or null. These topics might strike you as rather odd and specialized, so I need to explain why they are interesting and important for philosophy. Social epistemology, like individual epistemology, is infused with anthropocentric concepts – beliefs, propositional attitudes, intentional states, and many others. This orientation is outdated, if only because we must include scientific instruments and computers amongst the many sources and processors of knowledge. These things are sources not just of scientific knowledge, but of broader classes of knowledge. When a search engine is used to locate information on the Internet, the output of the automated algorithm underlying the engine can be the basis for the claim 'I know the top dozen in this list are among the most relevant sites for the search terms used'. In traditional epistemology, sources of knowledge need not possess knowledge themselves, although they may in the case of testimony, but we do speak of computers storing and processing knowledge as well as information, language that is not just metaphorical. Printed books contain knowledge and so do their on-line versions.

An important special case of network epistemology is a purely automated scientific network in which data is gathered by instruments and processed by computers without any intervention by humans. Some areas of science have already reached the automated stage – robotic astronomy, parts of experimental high energy physics, much of genomic analysis – and this trend is increasing. One reason to think through this scenario is that it is useful for a philosopher of

DOI: 10.3366/E1742360009000653

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press

science to consider the scientific enterprise from the perspective of an automaton. This can reveal the often subtle anthropocentric influences that are present even in apparently objective scientific activities. Sometimes these influences are not at all subtle, just unremarked. For example, despite its historical importance, empiricism as traditionally conceived is of no relevance to an automated science, not just because the devices involved do not possess anthropomorphic cognitive states but because the limits of the human perceptual apparatus are inappropriate constraints to put upon a scanning tunneling microscope or a computer assisted tomography imaging device. In a related way, the constraints of a priori human mathematical abilities do not matter to a computer checking a sub-case of the Kepler conjecture. It is revealing that whereas those who appeal to the traditional a priori in mathematics are usually willing to use arguments that appeal to in principle concepts of computability or deductive closure, empiricists have not been willing to be similarly generous.[2]

We should therefore be cautious in transferring results from traditional individualistic and social epistemology to automated science. Furthermore, the philosophical issues involved in computational science are very different from those that have been discussed in the philosophical literature on artificial intelligence. Contemporary scientific computers do not have conscious states and they are not constructed to mimic human modes of thought. They should not therefore be expected to have propositional attitudes. Scientific instruments do not have beliefs, nor do the computers that process data or perform theory based or agent based simulations. Of course, we could simply consider the traditional propositional attitudes as placeholders for whatever types of state the artifacts enter in the relevant epistemological contexts. So, within automated Bayesian inference, we could identify probability distributions with coherent degrees of belief, but then there is no point in using the belief talk; we may as well simply talk of probabilities. This is all to the good because the languages of belief, desire, hope, and so on are infused with subjective, anthropomorphic connotations that are best avoided.[3]

A PROPOSAL

We can, while rejecting this anthropomorphism, retain a connection with traditional epistemology. I propose that as a substitute for the traditional, pre-Gettier analysis of knowledge, we take this as our starting point:

> A computational device has knowledge of a system just in case the device possesses a true, evidentially supported model of the system.

The principal interest for computational purposes will be the modeling component of this definition, but it is worth addressing the evidential support and truth elements. Evidential support here is relatively unconstrained. It can be inductive support captured through an automated process of statistical testing, it could rely on formal inductive inference, it might employ Bayesian confirmation,

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press

or it may rest on corroboration in the Popperian sense, which in these contexts can include model selection by genetic algorithms. Other approaches are no doubt possible. Which of these we choose is of no particular importance within the context of this paper, although in particular situations there will be reasons for adopting one rather than others. The truth component will perhaps be problematical for some, not because they reject truth as a necessary condition for knowledge, but because they find the very idea of a model being true to be puzzling or inappropriate. We can address these concerns by noting that in particular cases, the model will provide an accurate representation of some parts of the system without providing a complete description of the system. More needs to be said when approximations and idealizations are used to construct the model, but we can deal with those using the position of selective realism that is presented in Humphreys (2004, section 3.8).

Using the definition above, we can make sense of claims of the form '**a** knows that system S is in state s', where **a** is any computational device that is using the model.[4] This immediately raises issues about whether **a** is a single machine or a network. To arrive at an answer we need to identify some special features of networks. In what follows, I shall avoid using the term 'social epistemology' and refer instead to 'network epistemology'. Whatever a sociology of machines might be like, it will not share many features of human communities. Here are some salient differences between network epistemology and social epistemology:

1. One reason for moving to social epistemology is the fact that humans are essentially bounded cognitive agents, having limited cognitive abilities that cannot be significantly expanded. Because of this, no single knower is or could be the locus of all scientific knowledge and there are thus epistemic benefits from using socially distributed knowledge. This is in sharp contrast to machines, which can have direct access to networked knowledge and within which the computational capacities of the nodes can be expanded. This immediately raises the issue of how we can individuate knowledge bearers within a network.

2. If we consider individual computers in the network as epistemic agents, then those agents have what human agents do not, direct thought transfer. The entire contents of a computational agent's knowledge base can be sent unchanged to another node in the network. Although it is true that this transfer is accomplished, with conventional computers, through linguistic representations and so one might argue that this is not essentially different from the use of written texts to transfer knowledge from one person to another, the situation is different in the two cases. With computers, exactly the same model and exactly the same knowledge can be used by more than one machine. There is no subjectivity of the kind that is common with human beliefs, where two humans rarely, if ever, have exactly the same set of beliefs. We can isolate components of models through modular programming so that holism is significantly reduced and we also know explicitly what the background beliefs are. The larger theories that serve as the construction base of the models, both scientific and mathematical, act as the explicitly stated

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press

machine equivalent of culturally shared background knowledge, employed by all participants in the computing network. There is no relativity of interpretation, no differential modes of abstraction, no semantic holism, and so on. Because of this direct transfer ability, it is possible to consider the entire network as a single knower with its knowledge dispositionally distributed across the computational nodes of the network.

Yet the issue is not straightforward. Consider a network in which one node has sufficient computing power and memory to perform all of the tasks that are currently distributed around the network. In this case, although the knowledge is distributed, it is not essentially distributed. Because any given computer can be part of an arbitrarily large network and conversely, all the theoretical knowledge in a given field could be collected in one machine, the **a** in '**a** knows that S is in state s' can in many cases be either a single machine or a network of computers, with the requirement that **a** has the computational resources to process the model of S.[5] I note that there is no need to consider the knowledge in these networks as being distributed in the way knowledge is in a neural net, where issues of subconceptual representation and distributed storage become central. The networks considered here have traditional non-connectionist architectures.

3. Models are often modular, in the sense that two models, even inconsistent models, can be run simultaneously on a computer without interactions by using parallel processors. Deductive closure conditions can be imposed within the separate models in the parallel machine with no negative consequences. (The inconsistent models do not jointly constitute the basis for knowledge, of course, since they cannot both be true.) Rationality conditions require that the entire belief set of a human be consistent because humans are considered to be unitary cognitive agents and modularization, in the form of split personality disorders, for example, is considered a sign of irrationality. But we do not feel the need to impose similar criteria on computers. The kind of belief holism that leads to extreme forms of subjectivity and conventionality about even individual propositional attitudes can be avoided because of this modular feature.

4. With an interpreted model, there is no issue about the level of conceptualization or representation that the computer is using, because it has access to no other. A cellular automaton operating with representations of the states of individual cells has no concept of a Turing machine, even though a Turing machine is present in some cellular automata with the appropriate initial conditions.[6]

HYBRID NETWORKS AND CONTENT

I'll begin by making a distinction between what I call *a hybrid scenario* and *a fully automated scenario*. In discussing these scenarios, I shall generalize the scope of networks to include traditional analog scientific instruments, such as radio telescopes, in addition to computers and computationally assisted instruments. The

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press

traditional instruments will be in causal contact with their targets, but their inputs or outputs may be transformed into a digital form that is accessible for computational processing. A hybrid scenario is then any situation within which some but not all scientific information is acquired by non-human agents and knowledge must be exchanged between human and non-human agents. We have been in the hybrid scenario since the late sixteenth century when scientific instruments first provided us with information about the world that we could not access without them. A key feature of the hybrid scenario is *the interface problem*: how knowledge can be effectively exchanged between human and non-human epistemic agents. Although this exchange is bi-directional, the most interesting aspect is how humans use and understand knowledge based on models that are used by instruments and computers. It is the more difficult issue because one can engineer artifacts specifically to deal with representations produced by humans, but there is only one human cognitive architecture to accommodate the reverse flow, that produced by the highly contingent and particular history of human evolution.

A fully automated scenario is any situation in which scientific knowledge is acquired or processed without any input from humans. The automated scenario is a little exotic for most people and for obvious reasons it is not easy to say in detail what it will be like. It ought not to seem so mysterious, because for those who consider human cognitive capacities to be the result of computational processes, there should be no difference in principle between networks of humans and networks of computers, or between either of those and hybrid networks with nodes including both types. But since the evidence for humans as computational devices is less convincing than it once seemed to be, I shall restrict myself to the hybrid scenario.

There are two philosophical areas that are relevant to hybrid scenarios. The first involves the distinction between conceptual and nonconceptual content and I shall argue that epistemological networks can deal with nonconceptual content. Concepts are taken here to be representations of properties, which may be atomic or compound. This position conforms to the view that concepts are sub-sentential features, although concepts are not limited to the sentential domain, because graphical and other kinds of representations have conceptual content. We can then adopt two distinct attitudes towards the interface problem. The first is to attempt, as humans, to understand the distinctively different kinds of concepts that are used by automata to interact with the world and with one another. These concepts will frequently be different from the ones that seem to humans to be natural representational devices, the reason being that what lends itself to computational efficiency is often different from what allows for representational ease. This is a standard problem in agent based models, where the basic rules of interaction are usually very simple, but the output from the model includes sophisticated patterns for which we humans may currently have no suitable predicates.

The second attitude is to allow that many parts of automated science will use nonconceptual content and that we must come to grips with what that entails.

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press

Nonconceptual content is usually associated with knowledge directly gained from perception, before it is organized by cognition or language. Taking Russell's knowledge by acquaintance and Kant's perceptual intuitions as examples of knowledge having nonconceptual content, knowledge by acquaintance can give us access to primitive reference and it suggests that instruments with direct causal connections to their source are capable of providing a primitive reference relation. Russell's comment 'I know the colour perfectly and completely when I see it and no further knowledge of it is even theoretically possible' (1912, 47) also applies to an instrument that is constructed to identify and isolate a single property, such as a given wavelength of light. Strawson (1959, 18) claimed that we can identify an object demonstratively if we 'can pick it out by sight or hearing or otherwise sensibly discriminate' that object. This point also generalizes to whatever properties and individuals a scientific instrument detects. For example, a biometric identification device demonstratively identifies me just in case it can distinguish me (by iris scanning technology, for example) from everyone else. This does not entail that we are committed to linguistic approaches when we deal with digital rather than analog devices because digitizing an analog image does not by itself introduce conceptual content. What I mean here is that digitization can simply preserve the spatial relations between elements of the image while using a discrete spatial representation rather than a continuous representation. Issues that arise when this digital image is coded into a binary language are different.

As I argued in Humphreys (2004, sections 2.7 and 5.2), instruments, including the human sensory organs, are property abstractors. For humans, the normal visual apparatus picks out at a distance the colour green from the temperature, kinetic energy, mass, and other properties of a jumping frog. We can thus happily commit ourselves to the position that abstraction is involved in perception, both for humans and for instruments. If a property has been isolated by an instrument, there is no need for a conceptual representation of that property in order to determine it; you get it for free. Because instruments can perform similar feats of property abstraction, the issue of conceptual versus nonconceptual content is not inescapably a psychological matter. In fact, concepts may turn out to be one of those features that persist only because of our anthropocentric biases and because of the need for a human/machine interface. For any but the most elementary computational models, however, humans can grasp only a small part of this structure, whereas in virtue of its superior inferential abilities, the machine has access to by far the largest part of that content.

For a computer, bits play the same role as immediate perceptions, sense data, or intuitions in humans. Computers do not need to have concepts in order to process machine language and for such machines, it is plausible that everything at the machine language or processing level is non-conceptual. The higher level programming languages are there only for the benefit of humans and the computational apparatus does not have access to those higher level concepts unless it is provided with a decompiler.[7]

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press

TESTIMONY

In the hybrid scenario, there are parallels between testimonial evidence from other people and our reliance on outputs from computational devices. In both cases we do not have direct access to the source of the evidence but must rely on the authority of an intermediary, in this case scientific instruments or computational devices. Both suffer from what I have elsewhere called epistemic opacity, the human inability to know in detail the processes that lead from the input of the device to the output. Even if we take into account the fact that humans are the result of natural selection whereas computers are the result of design, the details of the computational processes in both cases are largely hidden. And even if we take a naturalistic perspective on human cognition, we must still pay attention to the fact that human brains and most scientific computational devices have different architectures, languages, and degrees of accessibility.

So we can make a few standard classifications. Non-reductionists about testimony are willing to take testimonial evidence as a fundamental category of evidence, as long as there are no defeating conditions present, just as we take the evidence of our senses as epistemically basic, absent reason to do otherwise. In contrast, reductionists about testimony require positive reasons to accept testimony. In the hybrid scenario, it seems appropriate to take a reductionist attitude towards computational evidence because the reliability of the instrument must be taken into account. In Humphreys (2004) I argued that we need to know how an instrument works because we must ensure that no errors have been introduced. This is why the non-reductionist position is unacceptable: we need to know how the instrument works so that we have some positive reason to accept the evidence or we must have inductive evidence of the reliability of the instrument and ensure that no defeaters are present. But mere inductive reliability cannot be sufficient because instruments tend to work well within a specific domain of application and to become unreliable outside that domain. It is the knowledge of how the instrument processes the input/output stream that allows us to react appropriately when changes occur in the conditions under which the instrument operates.

A second distinction is between global reduction and local reduction. Global non-reduction is the view that all sources of testimony should be taken as trustworthy whereas local non-reduction is oriented towards a particular instance of testimony. Neither of these is appropriate for us. Instead we need to take a broad-based local reduction as the appropriate position for instruments because the reductionist position requires instrument-specific knowledge to determine whether a given source is reliable. The particulars of a given computer simulation are almost always relevant to this determination, despite the fact that the details of a particular simulation are rarely published. Yet once we have calibrated an instrument or simulation, any application of that device, within its limits of effectiveness, should be trustworthy. In Humphreys (2004) I used the riskiness of inferences from data

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press

to conclusion to motivate the epistemic priority of many instruments over humans. Sosa (2006) also uses the safety of an instrument's output as a gauge of the worth of its testimony.

To make a connection with what we discussed above, if the computational side of the interface has only nonconceptual content, it cannot count as testimony but it can count as evidence. This is because it is the content of the testimony that is transmitted and only propositional content, that is, conceptual content, can be testified to. Nevertheless, if testimony counts as a species of evidence, then both conceptual and non-conceptual content can play a role in our definition of knowledge given earlier.

REFERENCES

Humphreys, Paul. 2004. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method.* New York: Oxford University Press.

Russell, Bertrand. 1912. *The Problems of Philosophy*. London: Williams and Norgate.

Sosa, Ernest. 2006. "Knowledge: Instrumental and Testimonial." In J. Lackey and E. Sosa (eds.), *The Epistemology of Testimony*, pp. 116–23. Oxford: The Clarendon Press.

Strawson, Peter. 1959. *Individuals*. London: Methuen.

van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Oxford University Press.

NOTES

1 The paper read at the 2008 Leuven conference on computer simulations and social epistemology presented an agent based model for sub-maximizing agents. The paper published here was written specifically for this journal and addresses different, more philosophical, issues.

2 One prominent example is van Fraassen, who explicitly rejects going beyond the limits of current human perceptual abilities to ground scientific knowledge (van Fraassen 1980, 17). Empiricist philosophers of mathematics also frequently appeal to the limitations of human abilities.

3 This point does not support reductionist or eliminativist positions in the philosophy of (human) mind. The most plausible position to take about human mental states is that they are emergent.

4 I exclude here knowledge how and other kinds of knowledge that are dependent on particular embodiments. Viewed as a sufficient condition, the account can be applied to humans in addition to (artificial) computers.

5 I am assuming temporarily that we have a criterion for what counts as a single machine. There will be theoretical limits on the ability of a single machine to perform network-grade tasks, but we can set those aside here.

6 Finally, a significant segment of social epistemology seems to be motivated by political considerations from which I want to dissociate myself. Computers and accelerator detectors have no gender, race, capitalist yearnings, nothing.

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press

7 There is a meta-conceptual element to this in that it is humans who are conceiving the processes in terms of discrete bits and so on, but that is unavoidable. Also, there are issues about the role played by bytes, information packets, and so on in transmissions between nodes in the network, but these can be set aside here.

Paul Humphreys is Professor of Philosophy at the University of Virginia. His current principal areas of research are computational science and emergence. His most recent books are *Extending Ourselves: Computational Science, Empiricism and Scientific Method* (Oxford 2004) and *Emergence* (co-edited with Mark Bedau, MIT Press, 2008).

https://doi.org/10.3366/E1742360009000653 Published online by Cambridge University Press