What Scientists Know Is Not a Function of What Scientists Know

P. D. Magnus*†

There are two senses of 'what scientists know': An individual sense (the separate opinions of individual scientists) and a collective sense (the state of the discipline). The latter is what matters for policy and planning, but it is not something that can be directly observed or reported. A function can be defined to map individual judgments onto an aggregate judgment. I argue that such a function cannot effectively capture community opinion, especially in cases that matter to us.

In one sense, 'what scientists know' just means the claims that are the determination of our best science. Yet science is a collective enterprise; there are many scientists who have individual and disparate beliefs. So 'what scientists know', in another sense, means the omnibus composed of the epistemic state of scientist 1, the epistemic state of scientist 2, and so on, for the rest of the community. The phrase is ambiguous between a collective and an individual meaning.

If we consult a scientific expert, either because we want to plan policy or just because we are curious, we are typically interested in the collective sense. We want to know what our best current science has to say about the matter. And the expert we consult can differentiate the two senses, too. She can relate what she as a particular scientist knows (what she herself thinks, where her sympathies lie in controversies, etc.), but she can also take a step back from those commitments to give her sense of what the community consensus or dominant opinion is on the same matters. If it is simply curiosity that has led us to consult an expert, this may be enough.

*To contact the author, please write to: Department of Philosophy, University at Albany, SUNY, Albany, NY 12222; e-mail: pmagnus@fecundity.com.

†Thanks to John Milanese, Heather Douglas, and Jon Mandle for comments on various parts of this project.

Philosophy of Science, 80 (December 2013) pp. 840–849. 0031-8248/2013/8005-0006$10.00
Copyright 2013 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/673718 Published online by Cambridge University Press

When policy hangs on the judgment, however, we want more than just one expert's report on the state of the entire field.

This distinction between their personal commitments and the state of the field in their discipline is one that any scholar can make. If you think (as tradition has it) that only individuals can have beliefs in a strict sense, then take the expression 'opinion of the scientific community' as a façon de parler. If you think (as Lynn Hankinson Nelson does) that the community rather than the individual knows in a strict sense, then suitably reinterpret 'what an individual knows' in terms of belief (see Nelson 1990). The distinction I have in mind is neutral with respect to the metaphysics of social epistemology. The question is simply how we could use consultation with individuals to generate a composite, collective judgment.

Formal *judgment aggregation* offers a rigorous framework that seems to provide what we want. In the abstract, it defines a function that takes individual scientists' judgments as inputs and yields collective judgment as an output. This assumes that the collective judgment of the scientific community depends on the separate individual judgments of the scientists—that is, that what scientists know in the collective sense is a function of what scientists know in the individual sense.

Taking a recent proposal by Hartmann, Pigozzi, and Sprenger (2010) and Hartmann and Sprenger (2012) as an exemplar, I argue that judgment aggregation does a poor job of representing what scientists know in the collective sense. I survey several difficulties. The deepest stems from the fact that judgments of fact necessarily involve (perhaps implicit) value judgments. Where values and risks might be contentious, this entails that individual judgments cannot merely be inputs to a function. Judgment aggregation is not enough.

1. The Majority and Premise-Majority Rules. As a judgment aggregation procedure, one might naively survey scientists about factual matters and take any answer given by the majority of scientists to reflect the state of science. Of course, scientists would agree about a great many things that are simply not within their purview. Physicists would say that Sacramento is the capital of California, but that does not make it part of physics. So the survey should be confined to matters that are properly scientific. The survey must also include only legitimate scientists and exclude ignorant rabble. These restrictions are somewhat slippery, but let's accept them.

The naive procedure is a simple function from individual judgments to an aggregate judgment: return the judgment endorsed by a majority of the judges. Call this the *majority* rule.

The majority rule has the nice features that it treats every judge equally and that it does not bias the conclusion toward one judgment or another. Yet it suffers from what's called the *discursive dilemma*: it can lead to in-

https://doi.org/10.1086/673718 Published online by Cambridge University Press

consistent collective judgments, even if all the judges considered individually have consistent beliefs. In the following schematic example, there are three judges: Alice, Bob, and Charles. Each has the consistent beliefs on the matters P, Q, and $(P\&Q)$ indicated in the table below. The majority rule yields the inconsistent combination of affirming P and Q but denying $(P\&Q)$.

	P	Q	$(P\&Q)$
Alice	True	False	False
Bob	False	True	False
Charles	True	True	True
Majority	True	True	False

The nice features of majority rule seem like desiderata for a judgment aggregation rule, but avoiding the discursive dilemma is another such desideratum. A good deal of ink has been spilled specifying precisely the desiderata and proving that they are together inconsistent. However, even when it can be proven that a set of desiderata cannot be satisfied in all cases, they may still be jointly satisfied in some instances. Although the majority rule can lead to contradiction, it does not do so in every case. As a practical matter, we might begin by trying out a simple rule (like majority) and add sophistication only if the actual community has judgments like those in the schematic example.[1] Even so, more sophisticated rules would be needed for corner cases.

Hartmann et al. (2010) and Hartmann and Sprenger (2012) develop a judgment aggregation rule specifically to escape the discursive dilemma. Their procedure involves polling judges only regarding matters of independent evidence. For matters that are consequences of the evidence, the procedure derives consequences from the aggregated judgments. In the simple case given in the table above, for example, the procedure would affirm P and Q (because each is affirmed by a majority) and also $P\&Q$ (because it is a consequence of P and Q). Call this the *premise-majority* rule. When it can be applied, premise majority generates a consistent set of judgments.

There are several difficulties with premise majority, as a way of aggregating expert scientific opinion.[2] First, premise majority inevitably produces some determinate answer. As Brams, Kilgour, and Zwicker (1998) show, it is possible for a combination of separate elections to result in an

1. The strategy of adding complications only as necessary can be applied generally to decision problems. For example, intransitive preferences wreck dominance reasoning. Yet one might presumptively employ dominance reasoning until one actually faces a case in which there are intransitive preferences.

2. Since Hartmann et al. are thinking about the general problem of judgment aggregation, rather than the problem of expert elicitation, these are objections to the application of the rule rather than to the rule as such.

https://doi.org/10.1086/673718 Published online by Cambridge University Press

overall outcome that would not be affirmed by any of the voters. Moreover, a judge's inconsistency will necessarily be between some belief about evidence and some belief about the consequences of the evidence—since the evidence claims are stipulated to be independent—but premise majority does not query beliefs about consequences at all. So it will generate a consistent set of judgments even if many or all judges are inconsistent. As such, premise majority will generate determinate results even when the community is confused or fractured into competing camps. But, in considering scientific opinion, we certainly only want to say that there is something that 'scientists know' when there is a coherent scientific community.

Second, applying the rule requires a division between the judgments that are evidence and the ones that are conclusions. As Fabrizio Cariani notes, premise majority "requires us to isolate, for each issue, a distinguished set of logically independent premises" (2011, 28). He constructs a case involving three separate contentious claims and an agreed-on constraint, such that any two of the three claims logically determines the third. It would be arbitrary to treat two of the claims as evidence (and so suitable for polling) and the third as a consequence (and so fixed by inference). The premise-majority rule simply is not applicable in cases in which the line between premises and conclusions is so fluid. This difficulty leads Cariani to conclude only that premise majority will sometimes be inapplicable, so he suggests, "Different specific aggregation problems may call for different aggregation rules" (29). Yet the problem is especially acute for scientific judgment, because inference can be parsed at different levels. Individual measurements like '35° at 1:07 a.m.' are not the sort of thing that would appear in a scientific publication; individual data points are unrepeatable and not something about which you would query the whole community. Yet they do, of course, play a role in inference. At the same time, scientists may take things like the constancy of the speed of light to be evidence for a theory; the evidence here is itself an inference from experiments and observations.[3] Since we might treat the same claims as premises or conclusions, in different contexts, it is unclear what we would poll scientists about if we applied premise majority.

Third, premise majority is constructed for cases in which the conclusion is a deductive consequence of the premises. In science, this is almost never the case.[4] Scientific inference is ampliative, and there is uncertainty not only about which evidence statements to accept but also about which inferences ought to be made on their basis.

3. There are different labels for these different levels. Trevor Pinch (1985) calls them observations of differing externality. James Bogen and James Woodward (1988) distinguish data from phenomena.

4. I say "almost" because sufficiently strong background commitments can transform an ampliative inference into a deduction from phenomena. Of course, we accept equivalent inductive risk when we adopt the background commitments; cf. Magnus (2008).

https://doi.org/10.1086/673718 Published online by Cambridge University Press

One might avoid this difficulty by including inferential relations among the evidential judgments. To take a schematic case, judges could be asked about R and $(R \rightarrow S)$; if the majority affirms both, then premise majority yields an affirmative judgment for S. This reply reflects what John Norton (2003) calls a *material theory of induction*. The central idea is that most of the inductive risk in ampliative inferences is shouldered by conditional premises that Norton calls *material postulates*. Although material postulates are often of the form 'If R, then typically S' rather than the stricter $R \rightarrow S$, they nevertheless underwrite an inference from R to S. So one might think that asking about material postulates would allow us to use the premise-majority rule to aggregate scientific judgments.

This suggestion presumes that scientists can say, independently of everything else, whether the inference from R to S is appropriate. That is, it assumes that material postulates can be evaluated on a ballot apart from other matters. In the remainder of the article, I argue that this idealizes science too much. Whether a scientific inference is appropriate must be informed by more than just the particular evidence—the appropriate scientific conclusion depends (at least in many important cases) on the risks and values involved.

In the next section, I spell out more clearly the way in which inference can be entangled with values and risk. In the subsequent section, I return to it as a problem for premise majority. As we will see, it becomes a problem for more than just Hartmann et al.'s specific proposal. It is a problem for any formal judgment aggregation rule whatsoever.

2. The James-Rudner-Douglas Thesis. Here is a quick argument for the entanglement of judgment and values: there is a tension between different epistemic duties. The appropriate balance between these duties is a matter of value commitments rather than a matter of transcendent rationality. So making a judgment of fact necessarily depends on value commitments.

The argument goes back at least to William James, who puts the point this way: "*We must know the truth*; and *we must avoid error*—these are our first and great commandments as would-be knowers; but they are not two ways of stating an identical commandment, they are two separable laws" (1896/1948, 99). Although James has in mind personal matters of conscience (such as religious belief), Richard Rudner makes a similar argument for scientific judgment. Rudner argues that "the scientist must make the decision that the evidence is *sufficiently* strong . . . to warrant the acceptance of the hypothesis. Obviously our decision regarding the evidence and respecting how strong is 'strong enough,' is going to be a function of the *importance*, in the typically ethical sense, of making a mistake in accepting or rejecting the hypothesis" (1953, 2).

The tension between finding truth and avoiding falsehood can be expressed as the trade-off between two kinds of error. Any particular test in-

https://doi.org/10.1086/673718 Published online by Cambridge University Press

volves a trade-off between making the standards too permissive (and so mistakenly giving a positive answer) or making them too strict (and so mistakenly giving a negative answer). The former mistake is a *false positive* or *type I* error; the latter, a *false negative* or *type II* error. There is an inevitable trade-off between the risk of each mistake, and so there is a point at which the only way to reduce the risk of both is to collect more evidence and perform more tests. Yet the decision to do so is itself a practical as well as an epistemic decision. In any case, it leaves the realm of judgment aggregation—having more evidence would mean having different science, rather than discerning the best answer our current science has to a question. As such, values come into play. Heather Douglas puts the point this way: "Within the parameters of available resources and methods, some choices must be made, and that choice should weigh the costs of false positives versus false negatives. Weighing these costs legitimately involves social, ethical, and cognitive values" (2009, 104).

Plotting a curve through these nineteenth, twentieth, and twenty-first-century formulations, we call this the *James-Rudner-Douglas* or *JRD thesis*: anytime a scientist announces a judgment of fact, he or she is making a trade-off between the risk of different kinds of error. This balancing act depends on the costs of each kind of error, so scientific judgment involves assessments of the value of different outcomes.[5]

The standard objection to the thesis is that responsible scientists should not be making categorical judgments. They should never simply announce '*P*' (the objection says) but instead should say things like 'The available evidence justifies x% confidence in *P*'. This response fails to undercut the thesis because procedures for assigning confidence levels also involve a balance between different kinds of risk. This is clearest if the confidence is given as an interval, like $x \pm e$%. Error can be avoided, at the cost of precision, by making e very large. Yet a tremendous interval, although safe, is tantamount to no answer at all.

Justin Biddle and Eric Winsberg (2010) give a substantially more subtle reply to the standard objection. Regarding the specific case of climate modeling, Biddle and Winsberg show that scientists' estimates both of particular quantities and of confidence intervals depend on the histories of their models. For example, the results are different if scientists model ocean dynamics and then add a module for ice formation rather than vice versa. The history of a model reflects decisions about what was considered to be important enough to model first, and so it depends on prior value judgments.

But why should the JRD thesis have consequences for expert elicitation? After all, James does not apply it to empirical scientific matters. He is concerned with religious and personal matters, and he concludes merely

5. These three namesakes provide clear, prominent statements of the thesis, but of course they are not alone; e.g., see Lemons, Shrader-Frechette, and Cranor (1997).

https://doi.org/10.1086/673718 Published online by Cambridge University Press

that we should "respect one another's mental freedom" (1896/1948, 109). He does not apply it to scientific matters for which there is a community of legitimate experts.

Rudner, who does apply the thesis to empirical judgments, nevertheless hopes that the requisite values might themselves be objective. What we need, he concludes, is "a science of ethics" (1953, 6). Rudner calls this a "task of stupendous magnitude," but he is too optimistic. Searching for an objective ethics in order to resolve the weight of values and risks is a fool's errand. We would enter a vicious circle: the judgments of ethical science would need to be informed by the ethically correct values so as to properly balance inductive risks, but assurance that we have the correct values would only be available as the product of ethical science. One might invoke pragmatism and reflective equilibrium, but such invocations would not give Rudner final or utterly objective values. If responsible judgment aggregation were to wait on an utterly objective, scientific ethics, then it would wait forever.

Douglas accepts that the thesis matters for expert elicitation. So she considers the concrete question of how to determine the importance of the relevant dangers. She argues for an *analytic-deliberative process* that would include both scientists and stakeholders (2009, chap. 8). Such a process is required when the scientific question has a bearing on public policy, and there are further conditions that must obtain in order for such processes to be successful. For one, "policymakers [must be] fully committed to taking seriously the public input and advice they receive and to be guided by the results of such deliberation." For another, the public must be "engaged and manageable in size, so that stakeholders can be identified and involved" (166). Where there are too many stakeholders and scientists for direct interaction, there can still be vigorous public examination of the values involved. Rather than pretending that there is any all-purpose procedure, Douglas calls for "experiment with social mechanisms to achieve a robust dialog and potential consensus about values" (169). Where consensus is impossible, we can still try to elucidate and narrow the range of options. Douglas's approach is both a matter of policy (trying to increase trust in science, rather than alienating policy makers and stakeholders) and a matter of normative politics (claiming that stakeholders' values are ones that scientists should take into consideration). In cases in which these concerns are salient, saying what scientists know will depend on more than just the prior isolated judgments of scientists—it will depend, moreover, on facts about the actual communities of scientists, policy makers, and stakeholders.

Arguably, Douglas's concerns will not be salient in all cases. Some science is far removed from questions of policy. So the significance of the JRD thesis may depend on the question being asked.

https://doi.org/10.1086/673718 Published online by Cambridge University Press

3. Our Fallible Selves. I argued above that the premise-majority rule was inapplicable in many scientific contexts because it only worked for cases of deductive consequence. Formally, this worry could be resolved by asking scientists about which inferences would be justified; we poll them about claims like $(E \rightarrow H)$ at the same time as we poll them about E. The JRD thesis undercuts this formal trick. Where the judgment has consequences, the inference itself is an action under uncertainty. So the appropriate inference depends on the values at stake. Schematically, whether one should assent to $(E \rightarrow H)$ depends on the risks involved in inferring H from E. Concretely, questions of science that matter for policy are not entirely separable from questions of the policy implications.

If we merely poll scientists, then we will be accepting whatever judgments accord with their unstated values. We instead want the procedure to reflect the right values, which in a democratic society means including communities effected by the science. Importantly, this does not mean that stakeholders get to decide matters of fact themselves; they merely help determine how the risks involved in reaching a judgment should be weighed. Nor does it mean that politicized scientific questions should be answered by political means; climate scientists can confidently identify general trends and connections, even allowing for disagreement about the values involved. What it does mean is that scientists cannot provide an account that is value neutral in all its precise details.[6]

This is fatal to premise majority as a method of determining what scientists know collectively. Moreover, it is fatal to any judgment aggregation rule that treats judges merely as separate inputs to an algorithm. The problem extends to practical policies of expert elicitation, insofar as they are procedures for enacting judgment aggregation rules. Where there are important values at stake that scientists are not taking into account or where the value commitments of scientists are different from those of stakeholders, the current judgments of individual scientists cannot just be taken as givens.

So what should we do? It is worth distinguishing two kinds of cases.

First, in some cases, the problem could be ameliorated by an analytic-deliberative process that leads the scientists to consider the relevant values. However, the appropriate mechanisms are not ones that we can derive a priori. As Douglas argues, we need to experiment with different possibilities (2009, 169). There is not likely to be one universally applicable process. It will depend on facts about the communities involved. Moreover, the inference from social experiments in deliberation will itself be an inductive inference about a question that effects policy. So the inference de-

6. Douglas (2009, esp. chap. 6) provides an excellent discussion of how (what I have called) the JRD thesis is compatible with objectivity.

https://doi.org/10.1086/673718 Published online by Cambridge University Press

pends importantly on value judgments about the inductive risks involved, and that means an analytic-deliberative process will be required. It would be a mistake to hope, in parallel with Rudner's appeal to a science of ethics, for an objective set of procedural norms. How best to resolve metalevel judgment about experiments in social arrangements is as much a contingent matter as how to socially arrange object-level expert consultation.

Second, in other cases, it might be impossible for scientists to consider the relevant values by deliberation. Recall the example given by Biddle and Winsberg that the results of climate models depend on the sequence in which the modules were developed. Merely recognizing that different values would have led researchers to develop modules in a different order will not tell us what to believe because we do not know what different result that alternate pathway would have generated. What we want is some way of estimating the difference without having to start over and enact a different historical trajectory. It may be possible to do this at least in a qualitative way, for example, to estimate the direction or order of magnitude of various differences. But the ways of doing this will be local and contingent. They will probably also depend on prior value-laden choices.

In both kinds of cases, the solution is a turn to methods for assessing methodologies—for experimenting with analytic-deliberative procedures (in the former cases) or for evaluating the path dependence of object methodologies (in the latter). The JRD thesis can apply as much to these metamethdologies as to object methodologies. But we start with the best processes we can muster up now, and we try to improve them going forward. Minimally, we can say that future improvements should not elide the role of values, as formal judgment aggregation functions do, but explicitly accommodate it.

REFERENCES

Biddle, Justin, and Eric Winsberg. 2010. "Value Judgements and the Estimation of Uncertainty in Climate Modelling." In *New Waves in Philosophy of Science*, ed. P. D. Magnus and Jacob Busch, 172–97. Basingstoke: Macmillan.

Bogen, James, and James Woodward. 1988. "Saving the Phenomena." *Philosophy of Science* 97 (3): 303–52.

Brams, Steven J., D. Marc Kilgour, and William S. Zwicker. 1998. "The Paradox of Multiple Elections." *Social Choice and Welfare* 15 (2): 211–36.

Cariani, Fabrizio. 2011. "Judgment Aggregation." *Philosophy Compass* 6 (1): 22–32.

Douglas, Heather E. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.

Hartmann, Stephan, Gabriella Pigozzi, and Jan Sprenger. 2010. "Reliable Methods of Judgment Aggregation." *Journal for Logic and Computation* 20:603–17.

Hartmann, Stephan, and Jan Sprenger. 2012. "Judgment Aggregation and the Problem of Tracking the Truth." *Synthese* 187 (1): 209–21.

James, William. 1896/1948. "The Will to Believe." In *Essays in Pragmatism*, ed. Alburey Castell, 88–109. New York: Hafner.

https://doi.org/10.1086/673718 Published online by Cambridge University Press

Lemons, John, Kristin Shrader-Frechette, and Carl Cranor. 1997. "The Precautionary Principle: Scientific Uncertainty and Type I and Type II Errors." *Foundations of Science* 2 (2): 207–36.
Magnus, P. D. 2008. "Demonstrative Induction and the Skeleton of Inference." *International Studies in the Philosophy of Science* 22 (3): 303–15.
Nelson, Lynn Hankinson. 1990. *Who Knows: From Quine to a Feminist Empiricism*. Philadelphia: Temple University Press.
Norton, John D. 2003. "A Material Theory of Induction." *Philosophy of Science* 70 (4): 647–70.
Pinch, Trevor. 1985. "Towards an Analysis of Scientific Observation: The Externality and Evidential Significance of Observational Reports in Physics." *Social Studies of Science* 15:3–36.
Rudner, Richard. 1953. "The Scientist qua Scientist Makes Value Judgments." *Philosophy of Science* 20 (1): 1–6.

https://doi.org/10.1086/673718 Published online by Cambridge University Press