When Expert Disagreement Supports the Consensus

Finnur Dellsén

This is a preprint of an article whose final and definitive form will be published in the *Australasian Journal of Philosophy*. The *Australasian Journal of Philosophy* is available online at: http://www.tandf.co.uk/journals/.

Abstract

It is often suggested that disagreement among scientific experts is a reason not to trust those experts, even about matters on which they are in agreement. In direct opposition to this view, I argue here that the very fact that there is disagreement among experts on a given issue provides a positive reason for non-experts to trust that the experts really are justified in their attitudes towards consensus theories. I show how this line of thought can be spelled out in three distinct frameworks for non-deductive reasoning, viz. Bayesian Confirmation Theory, Inference to the Best Explanation, and Inferential Robustness Analysis.

1 Introduction

Since laypeople do not generally have access to relevant scientific evidence or the skills to analyze it, they are forced to rely on the testimony of scientific experts when evaluating scientific theories. However, one salient fact about scientific experts is that they frequently disagree amongst themselves. In public discourse, this fact is often taken as a reason not to trust experts, even on matters on which there is little or no expert disagreement. For example, climate scientists currently disagree about whether we are now experiencing a 'pause' in CO_2-induced global warming – and, if so, what explains that. This disagreement is frequently cited by so-called 'climate skeptics' as a reason to distrust climate science generally. In particular, this disagreement is appealed to as a reason to be skeptical of the nearly unanimously accepted theory that human influence is a significant contributing cause of the rise in mean global temperatures from the mid-20th century onwards.[1] Indeed, even those on the other side of the climate change debate have worried that the disagreement in question undermines the public's trust in climate science generally and anthropogenic climate change in particular.[2]

Moreover, there are reasons to think that scientists themselves take this sort of worry sufficiently seriously that they sometimes feel the need to mask their disagreements in order to gain or retain the public's trust. A case in point is John Beatty's (2006) historical study of the deliberations of a panel of expert geneticists, commissioned at the height of the nuclear age by the United States National Academy of Science to produce a report on the safe ranges of exposure to atomic radiation. The official report eventually issued, *The Biological Effects of Atomic Radiation* (1956), reported a consensus position on key issues such as the probable amount of mutation induced by different doses of radiation. However, the report also withheld significant information from the public about the extent to which these experts disagreed on a host of related issues, such as the degree of uncertainty involved and the manner in which they thought the results should be calculated. This information about the extent to which the panel members disagreed appears to have been deliberately withheld for fear of undermining the panel's epistemic authority with regard to the key issues they were requested to investigate.

In direct opposition to the line of thought that undergirds these responses, I shall argue that expert disagreement can provide non-experts with a (*pro tanto*) reason to believe that the experts are indeed trustworthy with regard to the claims on which they agree. Although my approach abstracts away from specific scientific controversies such as those mentioned above, it is worth noting that the argument spelled out below has impor-

[1] One widely read 'climate skeptic' blog argues that that climate scientists should not be trusted because they "can't even make up their minds about the temperature of the Earth, about whether or not there was a pause in global warming, let alone about what the climate might do next" (Worrall, 2016).

[2] A recent *Spiegel* article claimed that the controversy over a possible pause in CO2-induced climate change "sends confusing and mixed messages to the lay public" (Traufetter, 2009).

tant implications both for the way in which laypeople should evaluate the testimony of scientific experts, and for the manner in which such scientists should communicate their results to epistemically rational non-experts. For example, the current disagreement among climate scientists on the alleged ‘pause’ in CO_2-induced global warming may well provide both policy makers and the public with an additional reason to trust climate scientists regarding the issues on which they agree, including anthropogenic climate change. Looking at the issue from the climate scientists’ point of view, the argument of this paper suggests that climate scientists should not shy away from acknowledging scientific disagreements when communicating with rational policy makers and reasonable members of the general public.

2 Preliminaries

Let us say that there is a *consensus* on a proposition P within a group G just in case all or nearly all of the members of G have the same or a very similar positive doxastic attitude towards P. For example, if 96% of a group have credences above 0.9 in some proposition P, then for our purposes the group will count as exhibiting consensus on P.[3,4] Similarly, I will say that there is *disagreement* within a group with regard to P if and only if it is not the case that all or nearly all of those members of the group that have a doxastic attitude towards P at all have the same or very similar doxastic attitude towards P. Put differently, a group disagrees on P just in case there is no consensus on P among those members of the group who have doxastic attitudes towards P at all.[5] It will also be convenient to settle on some terminology: Let $con(T_1)$ be the proposition that there is consensus among the experts on the relevant scientific topic that T_1 is true;

[3] This definition is vague in at least two obvious ways, viz. in the phrases ‘all or nearly all’ and ‘same or very similar’. Arguably, the definition also inherits vagueness in appealing to what is presumably itself a vague notion of group membership. However, since nothing in what follows will turn on the exact extension the concept, our definition of ‘consensus’ is sufficiently precise for current purposes.

[4] It is worth noting that, by this definition, a group G does not count as being in a consensus on P if there is a significant subset of G that is indifferent about P.

[5] As Hawthorne and Srinivasan (2013, 11n) point out, it would be unnatural to say that S_1 and S_2 disagree on P if S_1 believes P and S_2 is agnostic about P. Similarly, it would be unnatural to say that a group disagrees about P if the only members of the group who do not have a specific doxastic attitude towards P have no such attitude at all, even if this is true of a quite significant proportion of the group’s members.

let $dis(T_2, ..., T_n)$ be the proposition that experts disagree about T_2-T_n; and let $jus(T_1)$ be the proposition that T_1 really is epistemically justified by the scientific evidence available to the experts.

My concern in what follows is with what a non-expert on a given topic, i.e. a layperson or novice, should infer from a combination of consensus and disagreement in a group of people that are perceived to be experts on that topic. I do not have any specific account of expertise in mind here; rather, what I say should be consistent with any plausible account of expertise (e.g. Goldman, 2001). However, two points are worth emphasizing in this regard: First, a non-expert will be assumed not to have access to, or the requisite skills to evaluate, the evidence or arguments on the basis of which the perceived experts form their opinions. Second, the group to which the non-experts are appealing will be assumed to consist of *perceived* experts – what Scholz (2009, 187) calls "experts in a subjective sense" as opposed to "experts in an objective sense". Thus it will be left open whether the members of the group in question really are experts (or deserve to be called by that name), and instead merely assumed that they are thought of as such by non-experts.

As my examples so far indicate, the focus of this paper will be on cases in which the consensus theory T_1 and the contested (i.e. disagreed-upon) theories T_2-T_n fall within the same domain of expertise. Indeed, for reasons that will become clearer below, my argument will not apply to cases in which the consensus theory T_1 and the contested theories T_2-T_n concern entirely different issues, roughly since that defeats any reason to think that experts who disagree on T_2-T_n would not agree on T_1 unless it really is justified.[6] Furthermore, I will only be concerned here with consensus and disagreement about factual issues for which it is possible to obtain epistemic justification, i.e. issues in which there is an in-principle discoverable fact of the matter as to whether the theories in question are true or false. So for example, if no moral claims are true and/or epistemically justified (as per some forms of moral anti-realism), then my argument below would not apply to consensus and disagreements about morality. Perhaps similarly, it has been argued that at least some mathematical hypotheses, such as

[6] Thus, for example, my argument does not imply that the fact that theoretical physicists widely disagree on philosophical or methodological issues – such as the existence of God, or whether Bayesianism provides the basis for a correct account of scientific reasoning – gives us any reason to think that consensus theories in physics are justified.

George Cantor's Continuum Hypothesis, are absolutely undecidable and perhaps even indeterminate (Clarke-Doane, 2013, 2014). If this kind of anti-realism about (some) mathematical hypotheses is correct, then my argument does not apply to those particular mathematical claims either.[7]

The discussion in the following sections aims to establish a relation of epistemic support between the fact that a group of experts disagrees on some theories and the claim that a consensus theory is epistemically justified by the evidence available to those experts. Note that this connection is not directly concerned with the *truth* of the theory on which there is expert consensus; rather, it it is concerned with the claim that the consensus theory is epistemically justified by the available evidence. The former is a first-order claim about some proposition P, while the latter is a second-order claim that some evidence E makes P epistemically justified. Of course, the two claims are closely related since it is at least plausible that the fact that a group of experts are epistemically justified in believing some claim is itself a reason to believe that the claim is true. However, it turns out to be surprisingly difficult to say exactly how evidence for second-order claims of this sort should affect our attitudes towards the corresponding first-order claims (see ?Fitelson, 2012; Feldman, 2014; Roche, 2014; Comesaña and Tal, 2015; Tal and Comesaña, 2015). To bracket this issue, I will not be assuming any specific epistemic connection between the first- and second-order claims.

In saying that disagreement among experts is a reason to believe that consensus theories are epistemically justified, I do not mean to deny that there may be other – possibly stronger – reasons to believe that such consensus theories are *not* justified. Put differently, I aim to establish that disagreement can be a *pro tanto* reason to believe that a consensus theory is justified, where a *pro tanto* reason for something is a reason that might be outweighed by other reasons against it. Furthermore, I want to acknowledge that this reason to believe that consensus theories are epistemically justified is *defeasible* in the sense that such a reason may be nullified or

[7] On the other hand, if and to the extent that mathematical hypotheses are decidable and determinate, my argument applies to mathematical claims as well as the kind of empirical claims I have used as examples so far. More generally, there is nothing in particular about the argument below that restricts its scope to empirical claims, except in so far as one takes an anti-realist attitude towards various sorts of non-empirical claims (e.g. moral or mathematical claims).

cancelled by the addition of new information. Thus, to be perfectly precise, what I will argue below is that in the absence of such further information, the fact that a group of scientific experts disagree on some theories is a *pro tanto* reason to believe that a theory on which they have reached a consensus is indeed epistemically justified.

A final note: In the next three sections, I shall be giving three versions of what I consider to be essentially the same argument. Each version appeals to a distinct epistemological framework for non-deductive reasoning, viz. (i) a 'Bayesian' framework that appeals to Bayesian Confirmation Theory, (ii) an 'Explanationist' framework that appeals to Inference to the Best Explanation, (ii) and a 'Robustness' framework that appeals to Inferential Robustness Analysis. Each framework has its advantages and limitations. For example, Bayesian Confirmation Theory is notoriously silent on how one should assign initial or 'prior' probabilities, while Inferential Robustness Analysis employs a rather vague notion of presuppositional diversity. This is not the place to choose between, or improve upon, these epistemological frameworks. Instead, I will work around each framework's limitations whenever possible and acknowledge where we have reached each framework's limits. However, the particular shortcomings of each of the three approaches is mitigated by the fact that essentially the same argument can be spelled out in all three frameworks; hence the argument's soundness is not held hostage to the feasibility of any single framework for non-deductive reasoning.

3 The Bayesian Approach

For our purposes, *Bayesian Confirmation Theory* (BCT) can be viewed as the conjunction of three claims. First, rational agents have, or can be represented as having, fine-grained beliefs known as *credences*. Second, for perfectly rational agents, these credences satisfy the Kolmogorov axioms of probability, which means that their credences at a given time in propositions can be represented as probabilities. Third, perfectly rational agents are required to to update their credences in light of new evidence in accordance with *Bayesian Conditionalization*, which holds that whenever you obtain some evidence E (and no other evidence), you should set your credence in H equal to the credence in H conditional on E that you had before you gained evidence E. Formally:

$$P'(H) = P(H|E) \tag{1}$$

where $P'(\cdot)$ the probability function you should adopt after obtaining E and $P(\cdot)$ is the probability distribution you had before. It follows immediately that obtaining E increases the credence a rational agent assigns to H just in case

$$P(H|E) > P(H) \tag{2}$$

According to Bayesians, this increase of rational credence in H by virtue of obtaining E explicates epistemic notions such as 'confirmation' and 'support'. That is, Bayesians hold that E *confirms* or *supports* H, relative to a probability distribution $P(\cdot)$, just in case (2) holds.

Notice that Bayesian Confirmation Theory (henceforth BCT) relativizes confirmation or support to a given probability function $P(\cdot)$. Now, some probability functions will license what we would intuitively think of as outrageous assignments of probabilities; relative to such probability functions, we will arrive at equally outrageous verdicts about whether some piece of evidence supports some theory. This brings us to a notorious limitation of BCT, which is that it is unclear how to distinguish acceptable from unacceptable probability functions.[8] Nothing in what I say below solves or ameliorates this problem; accordingly, I will not assume that probability functions must meet any general constraints. What I will do is identify a particular (seemingly reasonable) condition on probability functions for which the central claim of this paper – that expert disagreement supports consensus theories – is validated by BCT. The argument given in this section is thus a conditional one that *if* one's probability function satisfies a particular (seemingly reasonable) condition, *then* expert disagreement supports consensus theories by Bayesian lights.

Now, it is worth noting that our concern is not with whether a consensus among scientific experts about T_1, i.e. $con(T_1)$, by itself confirms that T_1 is justified. That is, our concern is not with whether the following inequality

[8] Indeed, for this reason, many Bayesians hold that BCT is merely a bare-bones framework for non-deductive reasoning that will need to be supplemented with substantive assumptions about acceptable probability functions (see, e.g., Howson, 2000; Strevens, 2004).

holds:

$$P\big(jus(T_1)|con(T_1)\big) > P\big(jus(T_1)\big) \tag{3}$$

Rather, we will be concerned with whether disagreement about T_2-T_n *boosts* or *increases* the confirmation conferred on $jus(T_1)$ by $con(T_1)$. This corresponds to the claim that $jus(T_1)$ is confirmed to a greater extent by $con(T_1)$ *and* $dis(T_2, ..., T_n)$ together than by $con(T_1)$ alone. Formally:

$$P\big(jus(T_1)|con(T_1) \wedge dis(T_2, ..., T_n)\big) > P\big(jus(T_1)|con(T_1)\big) \tag{4}$$

It turns out to be easily provable that (4) holds just in case experts are more likely to disagree on T_2-T_n if they all agree that a justified theory is true than if they all agree that an unjustified theory is true.[9] Formally, (4) holds just in case:

$$\begin{aligned} &P\big(dis(T_2, ..., T_n)|con(T_1) \wedge jus(T_1)\big) \\ &\qquad > P\big(dis(T_2, ..., T_n)|con(T_1) \wedge \neg jus(T_1)\big) \end{aligned} \tag{5}$$

In sum, then, we have that $dis(T_2, ..., T_n)$ boosts the confirmation of $jus(T_1)$ provided by $con(T_1)$ relative to any probability function $P(\cdot)$ that satisfies the condition that the probability of disagreement among experts on T_2-T_n is greater given that a consensus theory T_1 is is justified than it is given that the consensus theory is not justified.

What does this tell us about whether scientific disagreement provides

[9] *Proof:* Using a version of Bayes's Theorem, we write (4) as:

$$\frac{P\big(jus(T_1)|con(T_1)\big)P\big(dis(T_2, ..., T_n)|con(T_1) \wedge jus(T_1)\big)}{P\big(dis(T_2, ..., T_n)|con(T_1)\big)} > P\big(jus(T_1)|con(T_1)\big)$$

Rearranging, we get:

$$P\big(dis(T_2, ..., T_n)|con(T_1) \wedge jus(T_1)\big) > P\big(dis(T_2, ..., T_n)|con(T_1)\big)$$

By the Law of Total Probability, the right-hand can be rewritten as follows:

$$\begin{aligned} P\big(dis(T_2, ..., T_n)|con(T_1)\big) = {} & P\big(jus(T_1)|con(T_1)\big)P\big(dis(T_2, ..., T_n)|con(T_1) \wedge jus(T_1)\big) \\ & + P\big(\neg jus(T_1)|con(T_1)\big)P\big(dis(T_2, ..., T_n)|con(T_1) \wedge \neg jus(T_1)\big) \end{aligned}$$

Substituting and then rearranging, we get:

$$\begin{aligned} &\big(1 - P\big(jus(T_1)|con(T_1)\big)\big)P\big(dis(T_2, ..., T_n)|con(T_1) \wedge jus(T_1)\big) \\ &\qquad > P\big(\neg jus(T_1)|con(T_1)\big)P\big(dis(T_2, ..., T_n)|con(T_1) \wedge \neg jus(T_1)\big) \end{aligned}$$

Since $P\big(\neg jus(T_1)|con(T_1)\big) = 1 - P\big(jus(T_1)|con(T_1)\big)$, this is equivalent to (5).

us with an epistemic reason to trust scientists on related topics on which agree? On the one hand, one might feel that this purely formal result tells us very little, since there is nothing probabilistically incoherent about one's probability function violating (5). As we noted above, BCT is limited in that it is restricted to providing us with purely formal requirements upon rational credences. All we can say on this approach is that *if* a given probability function satisfies (5), then it follows that the disagreement over T_1 boosts the confirmation of $jus(T_1)$ provided by the corresponding consensus on T_1. Since BCT by itself does not forbid *any* probability assignment, this result is as much as we would have reasonably hoped to achieve within the Bayesian framework.

On the other hand, we can of course ask ourselves whether condition (5) is satisfied for the probability functions we would normally consider reasonable (just as we can ask ourselves whether it would be reasonable for someone's full beliefs to satisfy conditions that go beyond logical consistency). The following consideration suggests that the answer is positive: Presumably, it is more common for groups of scientific experts who have reached a consensus that an unjustified theory is true to also agree on other theories in their field *irrespective of whether those theories are justified.* After all, groups that agree on unjustified theories would presumably have a lower epistemic bar for reaching such a consensus as compared with groups that reach consensus on justified theories; and as a result of having a lower epistemic bar for reaching consensus, the former groups would more often have reached consensus on other theories as well, irrespective of whether these theories are justified. Differently put, scientific experts who have reached a consensus on an unjustified theory will less often disagree on other theories whose epistemic status is in question, since the fact that they reached a consensus on the unjustified theory suggests that they reach consensus relatively easily.[10]

This does suggest that a *reasonable* probability distribution – one that fits with reasonable background assumptions about the relative frequencies

[10] One could mention other reasons why experts who have reached a consensus on an unjustified theory will less often disagree on other theories. In particular, one could follow John Stuart Mill (1859) in arguing that disagreements are in general conducive to a more critical atmosphere within groups, which in turn makes unjustified theories less likely to be accepted and thus agreed upon in such groups (see also Foley, 2001; Moffett, 2007).

of expert disagreement in different circumstances – assigns a higher value to $P\big(dis(T_2,...,T_n)|con(T_1)\wedge jus(T_1)\big)$ than $P\big(dis(T_2,...,T_n)|con(T_1)\wedge\neg jus(T_1)\big)$, thus satisfying (5). Relative to such a probability distribution, it follows from the equivalence of (4) and (5) that disagreement over T_2-T_n boosts the confirmation of $jus(T_1)$ provided by the corresponding consensus on T_1. Of course, since there is nothing probabilistically incoherent about violating (5), not all probability distributions will count as reasonable by this criterion. Nevertheless, even the admittedly impoverished Bayesian framework has delivered a result that is at least somewhat informative since the fact that reasonable probability distributions satisfy (5) has been shown to imply that expert disagreement on T_2-T_n boosts the confirmation of $jus(T_1)$ provided by $con(T_1)$ *relative to* probability distributions of this kind.

4 The Explanationist Approach

The Explanationist approach to non-deductive reasoning is built on the rule of ampliative inference that Gilbert Harman (1965) famously called *Inference to the Best Explanation* (IBE). Roughly, IBE holds that one may infer H from E if H provides a better potential explanation of E than any rival explanatory theory. An explanation is said to be better than another to the extent that it does better on various *explanatory considerations.* Unfortunately, it is a contested issue what kind of factors should be construed as explanatory considerations, and one that I will avoid as far as possible in this paper. However, two factors appear universally accepted as explanatory considerations by proponents of IBE:

> *Simplicity*: All other things being equal, T_1 is a better explanation than T_2 if T_1 posits fewer entities and processes (or fewer kinds of entities and processes) than T_2.
>
> *Explanatory power*: All other things being equal, T_1 is a better explanation than T_2 if T_1 explains more facts (or more kind of facts) than T_2.

In order for my argument below to have the widest possible appeal among those who adhere to the Explanationist approach, I will only appeal to these two considerations in what follows.

Two important but sometimes overlooked points about IBE will need to be kept in mind in what follows. First, any remotely plausible version of IBE assumes a *Requirement of Total Evidence* to the effect that the evidence E from which one is inferring should include *all* of the agent's relevant evidence. One would clearly not be warranted in inferring a theory that provides a suboptimal explanation of one's total evidence even if it provides the best explanation of a proper part of that evidence. Second, many inferences that are commonly characterized as instances of IBE are *indirect* as opposed to *direct*. Whilst a direct IBE infers H from the fact that it best explains E, an indirect IBE infers H from the fact that it *follows from* some H' that best explains E. Incidentally, this feature of IBE was exploited in Gilbert Harman's (1965) seminal paper, in which he argued that enumerative inductions can be construed as instances of IBEs since the propositions of the form 'The next observed A will be B' follow from propositions such as 'All observed As are Bs', which in turn provide the explanation for evidential propositions of the form 'All observed As are Bs' (Harman, 1965, 91).

Now, recall that we are concerned with whether the fact that there is disagreement about T_2-T_n among a group of scientific experts provides a reason to believe that another theory on which there is consensus is epistemically justified by the experts' evidence. To answer this question using IBE, let us compare the epistemic situation of a non-expert who lacks the information about the scientific disagreement on T_2-T_n with a non-expert who has that information. Consider first the non-expert who is trying to evaluate whether she should infer that H is justified from just the fact that there is expert consensus on T_1. Since the only relevant evidence of such an agent is $con(T_1)$, she should infer $jus(T_1)$ by IBE just in case $con(T_1)$ is best explained by a theory that entails $jus(T_1)$.

However, it is not at all clear that, in the absence of any other relevant evidence, a consensus on some theory T_1 is best explained by such a theory. Consider the following two explanations[11] for $con(T_1)$:

> Critical Evaluation: Scientific experts form beliefs by examining the scientific evidence for a given theory critically and/or independently of each other; accordingly, they reached a consensus on T_1

[11] Here and in what follows, I am using the term 'explanation' as synonymous with 'potential explanation' rather than 'correct explanation'.

> because T_1 was overwhelmingly justified in light of that evidence.
>
> CROWD PSYCHOLOGY: Scientific experts form beliefs irrespective of the evidence by following the lead of their peers and/or the lead of some scientific authority; accordingly, they reached a consensus on T_1 because T_1 was believed by the relevant peers and/or authority.

The first of these explanations – CRITICAL EVALUATION – does entail $jus(T_1)$, while the second – CROWD PSYCHOLOGY – does not. The problem, however, is that it is hard to see how a dispute between proponents of each explanation could be settled if nothing but $con(T_1)$ is taken as relevant evidence. After all, each explanation accounts for the evidence at hand quite well, and neither explanation is clearly simpler than the other, so it would seem that the two explanations are on a par with regard to simplicity and explanatory power.[12]

It is worth noting that the dispute between proponents of explanations like CRITICAL EVALUATION and CROWD PSYCHOLOGY is not merely hypothetical. Anyone who is familiar with the debates sparked by Thomas Kuhn (1996) about the role of sociological influences on scientific research will know that explanations of the latter kind are frequently offered within sociological studies of science, most famously by Latour and Woolgar (1979) and various advocates of the 'strong programme' such as Barnes (1974) and Bloor (1991). While these views are often dismissed as being committed to an implausible form of relativism, the point here is that if we have nothing but $con(T_1)$ to go on when evaluating whether T_1 is justified, then it is hard to see on what grounds the explanation they favor, CROWD PSYCHOLOGY should be deemed worse than CRITICAL EVALUATION. After all, the agreement on T_1 by itself tells us nothing about how such an agreement was reached.

However, the situation changes if we add to the total evidence the fact that there is disagreement on other theories T_2-T_n. To see why, note that CRITICAL EVALUATION can easily explain $dis(T_2, ..., T_n)$ as due to the scientific evidence for T_2-T_n not being sufficiently univocal for scientists to reach identical conclusions concerning these theories. CROWD PSYCHOL-

[12] Of course, CRITICAL EVALUATION and CROWD PSYCHOLOGY (and its ilk) are not the only possible explanations for a consensus on T_1. However, I take it that these two explanations are by some distance the most plausible explanations available for such a consensus.

ogy, by contrast, is unable to explain why scientists would not also follow each other's leads (or that of an authority) on T_2-T_n as they are alleged to have done regarding T_1. Crowd Psychology, unlike Critical Evaluation, cannot invoke anything about the evidence for or against T_2-T_n since, according to the former explanation, such evidence has nothing to do with whether scientists reach consensus. In short, Crowd Psychology suffers in explanatory power compared to Critical Evaluation since the former fails to explain something that can easily be explained by the latter.

Of course, we can imagine a modification of Crowd Psychology that invokes some special reasons why T_2-T_n are such that the experts have failed to reach a consensus, presumably in terms of some peculiar sociological facts about the theories in question. The problem with this explanation (call it Crowd Psychology+) is that it suffers in simplicity. After all, this explanation will invoke some sociological particulars that have allegedly prevented the scientists in question from reaching a consensus on T_2-T_n as they did on T_1. Hence Crowd Psychology+ will inevitably be less simple than its predecessor (no matter what particular reasons it invokes to explain $dis(T_2, ..., T_n)$, and thus also less simple than Critical Evaluation. In sum, then, Crowd Psychology+ would make up for the predecessor's disadvantage with regard to explanatory power by sacrificing its previous parity with regard to simplicity.

The upshot is that while Critical Evaluation and Crowd Psychology are on a par both with regard to simplicity and explanatory power as explanations of $con(T_1)$ alone, the situation changes once we add $dis(T_2, ..., T_n)$ to the total evidence. At that point, Critical Evaluation emerges as a better explanation, since its most plausible competitor either suffers in simplicity (if it invokes special reasons to explain why there is no consensus on T_2-T_n) or explanatory power (if it does not). This helps make $jus(T_1)$ warranted by IBE in virtue of favoring an explanation from which $jus(T_1)$ follows – viz. Critical Evaluation – over a set of explanations from which it does not – viz. Crowd Psychology and its ilk. Thus, given the Explanationist approach to non-deductive reasoning, disagreement on T_2-T_n does indeed provide a defeasible, *pro tanto* reason to believe that a related consensus theory T_1 is justified.

5 The Robustness Approach

We turn finally to spelling out the argument by appealing to the scientific methodology known as *robustness analysis*, introduced by biologist Richard Levins (1966), and developed by William Wimsatt (1981), Michael Weisberg (2006) and others. The basic idea behind robustness analysis is to identify as real or true any result that holds under a variety of different assumptions about a given phenomenon. However, as many authors have pointed out, there are a variety of distinct methodologies discussed under this heading in both science and philosophy (Wimsatt, 1981, 2011; Orzack and Sober, 1993; Woodward, 2006; Calcott, 2011). Here we will be concerned with a type of robustness analysis that exploits what Woodward (2006) refers to as *inferential robustness*, viz. the stability of support for a given conclusion from some set of data under a variety of different assumptions. I will refer to this as *Inferential Robustness Analysis* (IRA).

Roughly following Woodward (2006, 219-220), this notion can be precisified as follows. Suppose we have a set of data D from which we hope to infer a conclusion C. Suppose also that D does not itself imply C, so that some additional assumption(s) are required for C to be inferred from D. Furthermore, suppose that a number of distinct possible assumptions A_1-A_m are available, but that our background knowledge does not provide us with any reliable means of evaluating which of these is correct.[13] If for all or most of these assumptions A_1-A_m, D and A_i together imply C, then C is said to be *inferentially robust* with respect to A_1-A_m given D. The idea behind the current type of robustness analysis is then that C's inferential robustness provides a strong reason for believing that C is true (given D) for a suitable set of assumptions A_1-A_m.

What counts as a suitable set of assumptions A_1-A_m? In an ideal case, we would know for certain that at least one of A_1-A_m is correct – a property that Woodward (2006, 221) refers to as "completeness". In such an ideal situation, it follows logically (and thus with certainty) that C is true if D is true. However, since this ideal case will rarely if ever materialize in any

[13] For simplicity's sake, each A_i will be taken to be a single proposition as opposed to a set of propositions. This involves no loss of generality since sets of multiple assumptions can be converted into a single proposition by taking their conjunction.

remotely realistic situations,[14] some weaker condition must be identified if IRA is to be applicable in such situations. To a first approximation, the weaker condition is that the assumptions A_1-A_m should be sufficiently *diverse* to reflect a spread of plausible possibilities and thus prevent the inference from depending on the specifics of those assumptions. To be more precise, since this kind of diversity is a matter of degree, we should say that the strength of an IRA depends on the degree of diversity present in the different assumptions A_1-A_m under which C is inferentially robust given D. Thus, as many authors have noted, the basic thought behind robustness analysis, including IRA, is that a result may be confirmed in virtue of its being robust under a sufficiently diverse set of assumptions about the phenomenon in question.[15]

We now apply this framework to the issue of what expert disagreement reveals about scientists' justification for consensus theories. Let us again compare a situation in which non-experts are aware of expert disagreement on theories related to the consensus theory and an otherwise identical situation in which they are not. Consider first the situation in which the non-experts are not aware of the the expert disagreement on other theories. The data, D, from which we are inferring in situations of this kind consists in the fact that there is consensus on T_1 among the relevant group of experts, $con(T_1)$, i.e. that all or nearly all of the experts agree that T_1 is true. The desired conclusion C is the proposition $jus(T_1)$, i.e. that T_1 really is epistemically justified by the evidence available to the relevant scientists. Finally, we let each assumption A_i represent the possibility that a given expert E_i evaluated the evidence for T_1 correctly so as to have an opinion on T_1 that really is justified by the available evidence. Given this setup, note that the desired conclusion, viz. $jus(T_1)$, does follow from all or nearly all conjunctions of D and A_i ($1 \leq i \leq m$), as required by IRA. However, in this situation, there is nothing to indicate that the assumptions A_1-A_m exhibit the kind of presuppositional diversity required for IRA to

[14] Indeed, completeness only seems satisfied when either (i) the assumptions A_1-A_m exhaust logical space, or (ii) some A_i is already known with certainty to be true. Neither situation seems to arise in scientific practice or everyday life.

[15] See in particular Wimsatt (1981), Weisberg (2006), and Schupbach (2016). Much more could be said about what sort of presuppositional diversity should be involved here and how exactly it correlates with the strength of an IRA. However, that task lies far beyond the of this paper (although see Schupbach (2016) for a detailed discussion of precisely this issue).

be applicable. After all, the relevant experts E_1-E_m may, for all we know, have reached their conclusion on T_1 by a similar or even identical process.

By contrast, consider an otherwise identical situation in which the information available to the non-experts include the fact that these experts disagree on T_2-T_n. This disagreement on theories related to T_1 strongly indicates that the experts in question have a variety of ways of evaluating evidence in their area of expertise, e.g. that they have different background assumptions and/or different standards for what counts as good evidence. So, in this latter type of situation, there is at least some relevant kind of diversity among the assumptions A_1-A_m. At the same time, the fact that these all or nearly all of these experts nevertheless agree on T_1 entails, for all or most assumptions A_1-A_m, that T_1 is epistemically justified. Thus IRA delivers the verdict that the additional information that experts disagree on T_2-T_n, $dis(T_2, ..., T_n)$, does indeed provide a reason to believe that T_1 is justified, $jus(T_1)$, given that these same experts have reached a consensus on T_1, $con(T_1)$.

Of course, *how much* support $jus(T_1)$ obtains in this way will depend on how diverse the assumptions really are – i.e. on how much diversity in the epistemic evaluations can be assumed to be in place among experts who disagree on other theories. However, the relevant point for our purposes is whether the expert disagreement on other theories provides *any reason at all* for non-experts to believe that T_1 is justified.[16] To see clearly that IRA gives an affirmative answer to this question, it is enough to highlight the contrast between the two situations described above with regard to presuppositional diversity. In the first situation, we have no reason at all to think that the experts used different or indeed non-identical methods for evaluating T_1 in light of the available evidence, and so we have no reason at all to think there is any diversity at all among the different assumptions A_1-A_n. In the second situation, by contrast, the fact that the experts evaluated T_2-T_n differently provides *at least some* reason to think the experts used different methods to evaluate T_1, which thus amounts to a boost in presuppositional diversity as compared to the previous situation. This shows unambiguously that expert disagreement helps support an inference from consensus to justification by IRA.

[16] Recall also, from section 2, that my thesis is only that such a disagreement would provide a *pro tanto* and *defeasible* reason to trust that the consensus theory is justified.

6 The Disagreement-to-Irrationality Objection

So far I have appealed to three distinct frameworks for non-deductive reasoning, arguing that in each framework expert disagreement provides non-experts with a (*pro tanto* and defeasible) reason to believe that the experts are justified in believing the theories on which they have reached a consensus. As noted in the introduction, this argument goes against a common response to expert disagreement according to which such disagreement undermines the trustworthiness of experts even with respect to consensus theories. In this section, I consider a possible argument for this response – what I call the *Disagreement-to-Irrationality Objection.* Although I contend that this argument is almost wholly without merit, I hope that discussing it sheds some light on why a position contrary to the one taken in this paper has seemed plausible to many authors, including a number of scientists and scientifically-interested laypeople (see section 1).

An informal statement of the the Disagreement-to-Irrationality Objection goes as follows:

> Since perfectly rational agents would reach the same or very similar conclusions on the basis of the same evidence, disagreement among experts on theories within their domain of expertise indicates that the experts formed the relevant beliefs at least somewhat irrationally. To the extent that this is so, we should not expect the beliefs formed by such experts to be epistemically justified by the available evidence, including the beliefs on which they are in agreement. Thus, disagreement among experts is a (*pro tanto*) reason to *distrust* those experts generally, including their judgment concerning theories on which there is consensus.[17]

[17] Note that I am construing the Disagreement-to-Irrationality Objection as purporting to provide a *pro tanto* reason to not to trust experts regarding consensus theories. One might think that this makes the objection compatible with the position argued for in this paper, since one could simultaneously have a *pro tanto* reason to ϕ and a *pro tanto* reason to $\neg\phi$. Unfortunately, however, this response will not work since in this case it would be *the very same thing* that would constitute a reason for an against trusting experts regarding consensus theories, viz. the fact that the same experts disagree on other theories.

This argument has some intuitive pull at first blush; however, I contend that its allure disappears when it is subjected to careful scrutiny. Let's first formalize the argument by putting it in standard logical form:

P1. That a group of experts disagree on T_2-T_n is a (*pro tanto*) reason to believe that they are epistemically irrational.

P2. That the agents in a group are epistemically irrational is a (*pro tanto*) reason to believe that theories on which the agents have reached a consensus, such as T_1, are not epistemically justified.

C. That a group of experts disagree on T_2-T_n is a (*pro tanto*) reason to believe that theories on which the agents have reached a consensus, such as T_1, are not epistemically justified. [From P1 and P2.]

My main criticism of this argument is that its appeal is due to an equivocation. Once this equivocation is resolved, it will become clear that there is no sound argument to be made here.

Before we get to my main criticism, let us note that P1 is only plausible in so far as it is credible that experts could not come to have conflicting doxastic attitudes towards the same propositions without some of them being epistemically irrational. This can be challenged in at least three ways: First of all, it should be uncontroversial that an otherwise rational agent may occasionally make mistakes, or performance errors, in her belief-forming process. Just as an excellent archer sometimes makes a poor shot, a highly rational agent sometimes forms doxastic attitudes that fail to be fully justified by her evidence. At most, then, P1 holds true in cases where the disagreement cannot plausibly be explained without positing a systematic defect in the agents' evaluations of the the relevant theories, e.g. in situations in which there is very widespread disagreement among the relevant experts. This restricts the scope of the objection.

Second, note that the experts in question could be forming their beliefs on the basis of different evidence, in which case the fact that they reach different conclusions from this evidence is precisely what one should expect from epistemically rational agents. Indeed, many – perhaps most – cases of expert disagreement are cases in which the experts have formed their beliefs on the basis of different sets of evidence. For example, expert disagreement about climate change is often between climate scientists who appeal to different kinds of climate models, or even between scientists from different branches of climate science, e.g. historical climatology and paleoclimatology respectively. Clearly, disagreement of this kind does not in any way indicate that either party is epistemically irrational. At most, then, the conclusion of the argument would hold true in the limited class of cases in which the experts in question form their doxastic attitudes based on the same evidence.

A third and related point is that the experts may possess different background beliefs, e.g. about the reliability of different kinds of evidence and about the relative weight one should put on different kinds of theoretical considerations. Such background beliefs will influence their reasoning about the theories in question in a perfectly rational manner, so that experts with different background beliefs will rationally come to different, but equally rational, conclusions in light of the same evidence. This shows that the scope of the argument must be restricted even further to those types of disagreements in which the parties have the same or very similar background beliefs.

Let me now turn to my main criticism. Suppose two or more parties disagree on some issues – and suppose, for the sake of the argument, that all the evidence is shared between the parties, that they have the same background beliefs, and that the disagreement is widespread. This clearly does not entail or even indicate that *all* of the disagreeing parties are epistemically irrational. At most, widespread disagreement indicates that *some* (perhaps even *most*, but certainly not *all*) of the disagreeing parties are epistemically irrational. Thus P1 must be precisified as follows:

P1′. That a group of experts disagree on T_2-T_n is a (*pro tanto*) reason to believe that *some* (but not all) of them are epistemically irrational.

The relevant question, then, is whether there is an argument from P1′ to the conclusion C.

Such an argument would have to employ the following as the second premise in place of P2:

P2′. That *some* (but not all) of the agents in a group are epistemically irrational is a (*pro tanto*) reason to believe that theories on which the agents have reached a consensus, such as T_1, are not epistemically justified.

However, this replacement for P2 is not plausible. To see why, note that it explicitly allows that *some* experts in the relevant group are epistemically rational. Accordingly, the premise provides no reason at all for believing that there are *no* experts who formed their opinion on T_1 and other consensus theories in an epistemically rational manner. However, barring massive performance-errors, any such subgroup of rational experts will have come to a positive conclusion about T_1 only if T_1 is indeed well-supported by available evidence. Since this subgroup constitutes part of the larger group which has reached a consensus on T_1, we would still have good reasons to believe that T_1 is epistemically justified by the available evidence. In sum, then, the fact that *some but not all* of the agents in some group are epistemically irrational will in general tell us very little, if indeed anything, about whether the theories on which they have reached a consensus are epistemically justified.

7 Conclusion

I have argued that non-experts can probe whether or not a theory that enjoys consensus among a group of experts is epistemically justified by considering whether the same group of experts disagree on other theories within their domain of expertise. Contrary to an assumption that is frequently made by scientists and laypeople alike, such disagreement should not undermine our trust in the consensus theory. On the contrary, disagreement of this kind provides us with a positive (although *pro tanto* and defeasible) reason to believe that the consensus theory is indeed epistemically justified. As noted in the introduction, one upshot of this is that disagreement is not necessarily something that scientists should strive to mask or conceal. In so far as scientists are communicating with rational agents who are acting in good faith, the best strategy for arguing that a consensus theory should be trusted may involve openly acknowledging that the same experts who reached the consensus disagree on a number of other theories within their domain of expertise.

References

Barnes, S. B. (1974). *Scientific Knowledge and Sociological Theory*. Routledge and Kegan Paul, Boston.

Beatty, J. (2006). Masking Disagreement among Experts. *Episteme*, 3:52–67.

Bloor, D. (1991). *Knowledge and Social Imagery*. Chicago University Press, Chicago, IL, 2nd edition.

Calcott, B. (2011). Wimsatt and the robustness family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings. *Biology and Philosophy*, 26:281–293.

Clarke-Doane, J. (2013). What is Absolute Undecidability? *Nous*, 47:467–481.

Clarke-Doane, J. (2014). Moral Epistemology: The Mathematics Analogy. *Nous*, 48:238–255.

Comesaña, J. and Tal, E. (2015). Evidence of Evidence is Evidence (Trivially). *Analysis*, 75:557–559.

Feldman, R. (2014). Evidence of Evidence is Evidence. In Matheson, J. and Vitz, R., editors, *The Ethics of Belief*, pages 284–300. Oxford University Press, Oxford.

Fitelson, B. (2012). Evidence of Evidence is Not (Nessarily) Evidence. *Analysis*, 72:85–88.

Foley, R. (2001). *Intellectual Trust in Oneself and Others*. Cambridge University Press, Cambridge.

Goldman, A. I. (2001). Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research*, 63:85–110.

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

Hawthorne, J. and Srinivasan, A. (2013). Disagreement Without Transparency: Some Bleak Thoughts. In Christensen, D. and Lackey, J., editors, *The Epistemology of Disagreement*, pages 9–30. Oxford University Press, Oxford.

Howson, C. (2000). *Hume's Problem: Induction and the Justification of Belief*. Clarendon Press, Oxford.

Kuhn, T. S. (1962/1996). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL, 3rd edition.

Latour, B. and Woolgar, S. (1979). *Laboratory Life: The Construction of Scientific Facts*. Sage Publications, Los Angeles, CA.

Levins, R. (1966). The Strategy of Model Building in Population Biology. In Sober, E., editor, *Conceptual Issues in Evolutionary Biology*, pages 18–27. MIT Press, Cambridge, MA, 1st edition.

Mill, J. S. (1956/1859). *On Liberty*. Bobbs-Merrill, Indianapolis.

Moffett, M. (2007). Reasonable Disagreement and Rational Group Inquiry. *Episteme*, 4:352–367.

Orzack, S. H. and Sober, E. (1993). A Critical Assessment of levins's 'The Strategy of Model Building in Population Biology (1966)'. *Quarterly Review of Biology*, 68:533–546.

Roche, W. (2014). Evidence of Evidence is Evidence under Screening Off. *Episteme*, 11:119–124.

Scholz, O. R. (2009). Experts: What They Are and How We Recognize Them – A Discussion of Alvin Goldman's Views. *Grazer Philosophische Studien*, 79:187–205.

Schupbach, J. (2016). Robustness Analysis as Explanatory Reasoning. *British Journal for the Philosophy of Science*, page To appear.

Strevens, M. (2004). Bayesian Confirmation Theory: Inductive Logic, or Mere Inductive Framework. *Synthese*, 141:365–379.

Tal, E. and Comesaña, J. (2015). Is Evidence of Evidence Evidence? *Nous*, DOI: 10.1111/nous.12101.

Traufetter, G. (2009). Stagnating temperatures: Climatologists baffled by global warming time-out. *Spiegel Online International*, URL: http://www.spiegel.de/international/world/stagnating-temperatures-climatologists-baffled-by-global-warming-time-out-a-662092.html (viewed April 8th, 2016).

U.S. National Academies of Science (1956). *The Biological Effects of Atomic*

Radiation. National Academy of Sciences – National Research Council, Washington, DC.

Weisberg, M. (2006). Robustness Analysis. *Philosophy of Science*, 73:730–742.

Wimsatt, W. C. (1981). Robustness, Reliability, and Overdetermination. In Brewer, M. and Collins, B., editors, *Scientific Inquiry and the Social Sciences*, pages 124–163. Jossey-Bass, San Francisco, CA.

Wimsatt, W. C. (2011). Robust re-engineering: a philosophical account? *Biology and Philosophy*, 26:295–303.

Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology*, 13:219–240.

Worrall, E. (2016). Shock discovery: Admitting climate uncertainty makes you more credible. *Watts Up With That*, URL: http://wattsupwiththat.com/2016/03/26/shock-discovery-admitting-climate-uncertainty-makes-you-more-credible/comment-page-1/ (viewed April 8th, 2016).