We Owe It to Others to Think for Ourselves

Finnur Dellsén

[To appear in J. Matheson and K. Lougheed (eds.), *Essays in Epistemic Autonomy*]

Abstract: We are often urged to figure things out for ourselves rather than to rely on other people's say-so, and thus be 'epistemically autonomous' in one sense of the term. But why? For almost any important question, there will be someone around you who is at least as well placed to answer it correctly. So why bother making up your own mind at all? I consider, and then reject, two 'egoistic' answers to this question according to which thinking for oneself is beneficial for the autonomous agent herself. I go on to suggest that the reason we should (sometimes) think for ourselves is that doing so (sometimes) increases the collective reliability of the epistemic community to which we belong. In many cases, this will do nothing at all to increase our own chances of forming correct beliefs. So, at least in this respect, the rationale for being epistemically autonomous is entirely 'altruistic'.

1. The Puzzle of Epistemic Autonomy

There are many questions that I am in no better position to answer than you are. I know that. And yet for many of those questions, I will not defer to your judgment. Doing so would certainly save time. Moreover, in many cases I'd be just as likely to get things right by listening to you rather than trying to figure out the answer for myself, since I know full well that you are as reliable regarding those topics – if not more so. Indeed, there are relatively few topics for which there wouldn't be *someone* that I could listen to whose judgment I take to be equally or more reliable than my own judgment. There are some exceptions, of course. For example, I know more than anyone else about most events in my own personal history, and I also arguably know more about some of my own mental states than anyone else. But for most other types of facts – most *public* facts, as we might call them – there will be someone I could consult that I consider no less reliable than myself.

So why do I bother making up my own mind about public facts at all? Why does any of us bother? The answer cannot be that it is worth the effort in terms of

reliably forming true beliefs, since *ex hypothesi* I am no more likely – and often less likely – to form correct beliefs by going it alone than by relying on others. Thus, in so far as I am thinking only of my own beliefs about public facts, and only of the likely truth-value of those beliefs, I apparently have no reason to figure things out for myself. And yet isn't there something objectionable about relying to such an extent on other people's say-so? Various Enlightenment thinkers, including Descartes (1985/1628), Locke (1975/1689) and Kant (1991/1784), appear to have thought so – an idea that is encoded in the widely-cited informal 'fallacy' known as 'appeal to authority'.

Let us call the tension between these two thoughts *the Puzzle of Epistemic Autonomy*. The puzzle, in short, is how to explain the value of critically evaluating claims for oneself in a world where there is almost always someone out there whose opinion is at least as likely to be correct as the opinion you would form on your own. Professional philosophers should take special interest in solving this puzzle, since their role in the educational system is often thought to consist largely in cultivating epistemic autonomy in their students (e.g., Nussbaum 2017). Indeed, an entire genre of philosophical education – critical thinking – is quite explicitly designed to do precisely that. For example, a popular critical thinking textbook announces in the preface that "[c]itizens who think for themselves, rather than uncritically ingesting what their leaders and others with power tell them, are the absolutely necessary ingredient of a society that is to remain free" (Cavender and Kahane 2009: xiv).

In previous work (Dellsén 2020), I have begun to develop what I would now characterize as a *qualified, altruistic* solution to this puzzle: We should be epistemically autonomous because, and to the extent that, it makes the consensus positions of experts more reliable, which in turn benefits the community as a whole by enabling laypeople to rely on such expert consensuses. This solution is 'qualified' because it does not imply that epistemic autonomy is always valuable – rather, it is valuable only in cases where one is among those who might be consulted on the relevant issue. And this solution is 'altruistic' in the sense that the value of someone being epistemically autonomous is not taken to consist exclusively in the way it effects their own epistemic situation. On the contrary, it is my contention that any non-altruistic – i.e., *egoistic* – explanation for why we should think for ourselves will be unable to capture what's distinctively valuable about epistemic autonomy.

In this chapter, my main aim is to scrutinize two egoistic solutions to the Puzzle of Epistemic Autonomy. These solutions are importantly different in that one of them

implies that epistemic autonomy is *directly* valuable for the autonomous agent, while the other implies that epistemic autonomy is valuable to the agent only *indirectly*, through positively affecting the community of which the agent is a part. To mark this distinction, I distinguish between *directly egoistic* and *indirectly egoistic* solutions to the Puzzle of Epistemic Autonomy, and then distinguish those from *altruistic* solutions, in section 2. I then examine, in section 3, a directly egoistic solution which appeals to the value of understanding, arguing that this fails to solve the Puzzle of Epistemic Autonomy. In section 4, I similarly argue that an indirectly egoistic solution, which appeals to the apparent value of disagreement, fails to solve the problem as well. In section 5, I conclude by outlining my own altruistic solution to the puzzle.

2. Direct Egoism, Indirect Egoism, and Altruism

What are the possible solutions to the Puzzle of Epistemic Autonomy? Let me start by restating the puzzle somewhat more precisely:

> *The Puzzle of Epistemic Autonomy:* In situations where an agent S could effortlessly access opinions on a particular set of propositions $\{P_i\}$ that she herself takes to be at least as reliable as her own, what epistemic reasons (if any) are there for S to critically evaluate $\{P_i\}$ for herself rather than adopting these opinions as her own?

One type of response to this puzzle is to reject that it has any solution. On this view, there are never any epistemic reasons to critically evaluate something when one has easy access to equally or more reliable opinions on that issue. Although defending such a response arguably puts a professional philosopher in an uncomfortable position, at least a handful of philosophers – Foley (2001), Huemer (2005), Zagzebski (2007), and Constantin and Grundmann (2018) – have defended positions of this kind. For example, Zagzebski (2007) argues that truth-seeking agents who prefer reaching their own conclusion as opposed to deferring to equally or more trustworthy peers would be *incoherent* since they must implicitly view themselves as more trustworthy even if they have no reason to do so.[1] A somewhat more moderate position is defended by Huemer (2005: 524-525), who argues that critical thinking is epistemically

[1] I discuss this argument in more detail in Dellsén 2020: 347-350.

irresponsible because it is simply less reliable than alternative methods for belief-formation, including appealing to experts.[2]

While these responses are admirably brave in carrying a particular type of argument to its apparent logical conclusion, they leave a huge explanatory gap: If epistemic autonomy really is incoherent or irresponsible, why have we evolved – biologically as well as culturally – to think things through for ourselves, and to view it as admirable in others? Thus, a more satisfying response would attempt to *solve* the Puzzle of Epistemic Autonomy rather than dismiss it in the aforementioned ways. In particular, a solution to the Puzzle of Epistemic Autonomy, with regard to a particular set of propositions $\{P_i\}$, would describe the epistemic reasons there are for S to critically evaluate $\{P_i\}$ in situations of this kind. On one way of carving up logical space, the types of epistemic reasons S might have for doing this can be divided into three broad categories:

> *Directly egoistic reasons*: S has a *directly egoistic reason* to critically evaluate $\{P_i\}$ for themselves if and only if doing so *(partially) constitutes* S's being in a superior epistemic state with regard to $\{P_i\}$.
>
> *Indirectly egoistic reasons*: S has an *indirectly egoistic reason* to critically evaluate $\{P_i\}$ for themselves if and only if doing so *(partially) causes* S to be in a superior epistemic state with regard to $\{P_i\}$.
>
> *Altruistic reasons*: S has an *altruistic reason* to critically evaluate $\{P_i\}$ for themselves if and only if doing so *(partially) causes other agents* to be in a superior epistemic state with regard to $\{P_i\}$.[3]

Here, a 'superior epistemic state' is any kind of state that carries more epistemic value – broadly defined – than another. What types of states have epistemic value is to be determined by the solution in question, but plausible candidates include *truth*, *accuracy*, *reliability*, *justification*, *knowledge*, and *understanding*.

In fact, one of the two proposals I examine below is that an agent has epistemic reasons to think for themselves because, and in so far as, doing so is partly constitutive of *understanding*. The other proposal I will consider is that an agent's thinking for

[2] Huemer also appears to endorse the claim that epistemic autonomy is incoherent (see Huemer 2005: 525-526).

[3] There is no category here corresponding to *directly altruistic reasons*, since such reasons would entail that S's critical evaluation of $\{P_i\}$ could somehow *constitute* the other agent's being in a superior epistemic state with regard to $\{P_i\}$. I suppose this could be palatable to certain hardline externalists, but I won't consider this option further here.

herself partially causes disagreements to arise in her epistemic community, which in turn is conducive to the agent's search for truth. These two proposals exemplify solutions to the Puzzle of Epistemic Autonomy that appeal, respectively, to directly and indirectly egoistic reasons for being epistemically autonomous. As noted, I will argue that neither solution is successful. However, since these solutions clearly do not exhaust the logical space of (directly or indirectly) egoistic reasons for epistemic autonomy, it may well be possible to develop an entirely egoistic solution to the Puzzle of Epistemic Autonomy for all that I have to say below. What my discussion does suggest, however, is that developing a plausible egoistic solution is more difficult than it may seem at first blush. And that, in turn, lends credibility to altruistic solutions to the Puzzle of Epistemic Autonomy, i.e. that an agent has epistemic reasons to think for themselves because, and in so far as, doing so partially causes *other agents* to be in superior epistemic states.

It should be said that one obvious type of reason for being epistemically autonomous with regard to some set of propositions $\{P_i\}$ is to develop the reasoning skills that one can then apply in other cases, i.e. with regard to another set of propositions $\{Q_i\}$. For example, a philosophy graduate student narrowly interested in metaphysics might have reason to think critically about normative ethics purely because they hope to develop general philosophical skills that they can later apply in their research in metaphysics. In such cases, the value of being epistemically autonomous regarding $\{P_i\}$ is to become better at being epistemically autonomous regarding $\{Q_i\}$. Clearly, this type of consideration in favor of epistemic autonomy only pushes the problem one step back, however, since we must now figure out what is valuable about being epistemically autonomous regarding $\{Q_i\}$.

3. Understanding and Direct Egoism

Several epistemologists have floated the idea that *understanding*, in contrast to more familiar epistemic states like knowledge and true belief, cannot be transmitted via testimony. You can inherit other people's knowledge or true belief that P through being told that P, but you cannot in the same way inherit other people's understanding. According to Zagzebski (2008: 146), "understanding cannot be given to another person at all except in the indirect sense that a good teacher can sometimes recreate the conditions that produce understanding in hopes that the student will acquire it also." Zagzebski's view is fairly typical among early pioneers of

understanding in epistemology (see, e.g., Hills 2009, Pritchard 2010). At least at first blush, it is also plausible, for it does seem that truly *understanding* any relatively complex phenomenon, e.g. the spread of an infectious disease like COVID-19, involves a distinct cognitive effort that is not required for simply believing or knowing what someone tells you about the disease.[4]

Supposing that Zagzebski's no-understanding-through-testimony view is correct, this might seem to undergird a solution to the Puzzle of Epistemic Autonomy. For if understanding cannot be transmitted via testimony, then agents must achieve understanding for themselves if they are to achieve it at all. It seems to follow that understanding requires agents to think for themselves – that they must themselves evaluate the propositions on the basis of which they understand. If understanding is distinctively epistemically valuable in some way – a very common sentiment that is both intuitive (Pritchard 2009, Elgin 2017) and seems to lie at the heart of our conception of scientific progress (Dellsén 2016, Potochnik 2017; see also Bird 2007: 84) – it follows that agents have a direct egoistic reason to exercise epistemic autonomy. In slogan form, *we should think for ourselves because we can only understand for ourselves*.

The trouble with this suggestion, however, emerges when we start to specify what exactly it is about understanding that would make it impossible to transmit it directly via testimony. Suppose I am trying to understand the rapid spread of COVID-19 in the spring of 2020. My potential understanding is based on several bits of information, including not only particular facts like its basic reproduction number R_0 at different times and locations, but also more general claims like the SIR model for spread of infectious diseases. All these bits of information can clearly be transmitted via testimony; otherwise, you and I would not know about them. However, once I have obtained this information, I do not thereby understand the spread of COVID-19. Roughly speaking, this is because I may not 'see' how these bits of information fit

[4] As Boyd (2017) points out, this might not be true of absolutely all cases of understanding, since achieving understanding based on various bits of information sometimes involves no real effort at all. For example, if you know that your friend has an early class to teach, and that your friend knows that there is construction on the road to the university, then it involves little effort to understand why your friend left their house especially early this morning (Boyd 2017: 14).

Malfatti (2020; see also Malfatti 2019) argues that understanding can be transmitted through testimony without significantly more effort on the receiver's end than in cases of knowledge through testimony. However, Malfetti's argument would not undermine the current solution to the Puzzle of Epistemic Autonomy since it would, in conjunction with the argument below, simply show that there is a wider range of epistemic states (viz. understanding *and knowledge*) that epistemic autonomy would essentially contribute to producing.

together, e.g. how R_0 depends on assumptions about the number of susceptibles (S), infected (I), recovered (R) in the relevant population.

What would be missing in such a case is what philosophers have come to call 'grasping' (e.g., Kvanvig 2003; Grimm 2006; Khalifa 2013; Strevens 2013). *Grasping* is thus the psychological component of understanding that goes beyond merely representing various bits of information and involves somehow representing these as a coherent whole. My own preferred view of grasping is that it consists in using these bits of information to construct a particular kind of *model* of the understood phenomenon – a model which represents whether and how each aspect of the phenomenon *depends*, e.g. causally or constitutively, on other aspects of the phenomenon (Dellsén forthcoming). Others construe grasping as involving a particular kind of ability (Wilkenfeld 2013), intelligibility (de Regt 2017), cognitive control (Hills 2016), or phenomenology (Bourget 2017). On any of these accounts of grasping, it involves something that one arguably needs to achieve for oneself once one has obtained the various bits of relevant information.[5]

The key question, then, is whether epistemic autonomy is essential for grasping. This might seem to be so at first blush since grasping is something one needs to do for oneself and epistemic autonomy is naturally described as 'thinking for oneself'. However, the connection here is illusory: Epistemic autonomy concerns not some vague and general idea that one must use one's own mental faculties; consulting other people also requires one to use mental faculties. Rather, epistemic autonomy consists in *critically evaluating* particular propositions as (probably) true or false, and making up one's mind on that basis. (Hence the 'epistemic' in 'epistemic autonomy'.) So if grasping constitutively involved being epistemically autonomous, grasping would have to be partly constituted by this process of making up one's mind on the basis of a critical evaluation of the relevant propositions.

However, none of the extant accounts of grasping take it to involve such a critical evaluation. For example, having cognitive control of a representation (Hills),

[5] In the course of arguing against 'lucky' understanding, Khalifa (2013; see also Khalifa 2017, ch. 7) proposes a partial account of grasping according to which *reliable evaluation of alternative explanations* is necessary for grasping. Khalifa's argument for this claim rests on the assumption that reliable evaluation of alternative explanations is necessary for understanding – an assumption that I think should be rejected (roughly on the same grounds that I reject justification requirements on understanding – see Dellsén 2017, 2018a). Furthermore, even if reliable evaluation of alternative explanations were necessary for grasping, this would be irrelevant for our current purposes since one way – indeed, the most reliable way in most cases – to evaluate explanations would be to consult (other) experts rather than to attempt to evaluate them for oneself.

or being in a certain phenomenological state (Bourget), do not require agents to critically evaluate the relevant. Nor is there any pre-theoretical reason to think that grasping requires critical evaluation. To see this clearly, consider again my understanding of the spread of COVID-19 in the spring of 2020. I indicated earlier that a (high degree of) understanding would require 'seeing', i.e., grasping, how various bits of information hang together – e.g., how R_0 depends on assumptions (in an SIR-model) about the number of susceptibles (S), infected (I), recovered (R) in the relevant population. This does not require epistemic autonomy in the relevant sense, since one can grasp such connections without making up one's mind on the basis of a critical evaluation. For example, I could come grasp this connection by being taught how to derive R_0 from S, I and R (and other assumptions), without in any way critically evaluating my teacher's instructions. Indeed, I could be so dependent on my teacher that there is absolutely no chance whatsoever that my grasp differs in any way from my teacher's. Hence I could come to understand the spread of COVID-19 without in any way making up my mind on the basis of a critical evaluation, i.e. without excercising epistemic autonomy.

I conclude that, since grasping does not require critical evaluation, agents do not need to exercise epistemic autonomy in order to achieve understanding. To be sure, grasping, and thus understanding, does require agents to engage in 'thinking' in the broadest sense of the term, but it does not require agents to engage in critical epistemic evaluations. Hence it does not require epistemic autonomy in the sense that the Puzzle of Epistemic Autonomy is concerned with. So the puzzle remains: What exactly is the value of critically evaluating various propositions for oneself, e.g. as opposed to merely organizing these propositions in an understanding-constituting way, when there are other people available who are in at least as good a position to perform that evaluation as you are yourself?

4. Disagreement and Indirect Egoism

The other 'egoistic' solution to the Puzzle of Epistemic Autonomy that I will consider is inspired by John Stuart Mill's defense of free speech. In *On Liberty* (1859), Mill famously argued that it is an essential part of rational inquiry to subject one's positions to counterarguments from those with whom one disagrees (Mill 1859: 45). Richard Foley interprets Mill as arguing that "disagreements encourage further evidence gathering and, thus, are in the long-term conducive to the search for truth" (Foley

2001: 124).[6] A similar thesis is advanced by Helen De Cruz and Johan De Smedt (2013), albeit without reference to Mill. Based on a case study concerning the taxonomic status of a recently discovered early human species called *Homo floresiensis*, De Cruz and De Smedt argue that scientific disagreement "is valuable because it brings about an increase in relevant evidence, a re-evaluation of existing evidence and assumptions, and a decrease in confirmation bias." (de Cruz and de Smedt 2013: 176). Presumably, these effects are themselves valuable because, and in so far as, they are conducive to the search for truth (as per Foley's suggestion).

The Millian point, then, is that disagreements tend to increase the reliability of individual experts in the long term, e.g. in virtue of encouraging them to gather more evidence than they otherwise would. If correct, this point would seem to show that epistemic autonomy is valuable, other things being equal, since disagreements would seemingly only arise amongst epistemically autonomous agents. Indeed, Foley immediately adds to the passage quoted above: "A corollary of this thesis [that diversity of opinion is in general preferable to unanimity] is that anything that discourages disagreements is potentially dangerous" (Foley 2001: 124). The idea, then, is that an agent S has an epistemic reason to critically evaluate $\{P_i\}$ because, and in so far as, doing so increases disagreement or dissent within the community, which in turn would increase S's chances of getting at the truth regarding $\{P_i\}$ in the long run. This would be an *indirectly egoistic reason* to be epistemically autonomous since critically evaluating $\{P_i\}$ would partially *cause*, rather than *constitute*, S's being in a superior epistemic state with regard to $\{P_i\}$, where the 'superior epistemic state' in this case consists in forming more reliably true beliefs about $\{P_i\}$ than S would otherwise do.

Note that this 'Millian' solution to the Puzzle of Epistemic Autonomy consists of two distinct claims. The first claim is that epistemic autonomy increases disagreement within a community. I take this to be a causal claim to the effect that epistemic autonomy either causes disagreement to arise, or causes there to be more disagreement than there would otherwise be. The second claim is that increasing disagreement within a community is conducive to each individual's search for truth. Again, I take this to be a causal claim to the effect that increasing disagreement causes the disagreeing agents to form true beliefs more reliably. So there are two distinct

[6] See also Matheson 2015 and Lougheed 2020, which is inspired by Elgin 2010.

causal claims contained within this solution to the Puzzle of Epistemic Autonomy. Let us consider these in turn.

The first claim holds that epistemic autonomy regarding some set of propositions $\{P_i\}$ increases the extent to which the community disagrees about $\{P_i\}$. There is a grain of truth in this. If a community is such that *none* of its members exhibit *any* epistemic autonomy at all, then there can be no variation in what the community takes to be true, and thus no disagreement. After all, a completely non-autonomous agent's views will be fully determined by the views of other members of their community, so that any two (or more) members of the same community will necessarily accept exactly the same things. Hence it is true that exhibiting *some* degree of epistemic autonomy is a precondition for disagreements to arise in the first place. However, in order to fully solve the Puzzle of Epistemic Autonomy we would need to a account for the value not just of *some* agents exhibited *some* autonomy but also of greater degrees of autonomy exhibited by a greater number of agents. Since the current solution connects the value of epistemic autonomy with the putative value of community disagreement, there would thus have to be some more general connection between the extent to which agents exhibit epistemic autonomy and the extent to which they disagree.

However, it is not hard to see that there is no such general connection between autonomy and disagreement. Suppose that a community of *n* individuals contains just two agents, X_1 and X_2, that are fully epistemically autonomous with regard to some set of propositions $\{P_i\}$, and that all other agents in the group, $X_3,...,X_n$, are fully non-autonomous with regard to $\{P_i\}$.[7] Specifically, suppose that half of the non-autonomous agents – $X_3,X_5,\ldots$ – automatically accept what X_1 accepts regarding a particular proposition P_i in $\{P_i\}$, and that the remaining agents – $X_4,X_6,...$ – automatically accept what X_2 accepts regarding P_i. In this sense, X_1 and X_2 can be said to be *authorities* on P_i for $X_3,X_5,\ldots$ and $X_4,X_6,...$ respectively. Now suppose that X_1 and X_2 disagree about P_i, i.e. that X_1 accepts P_i while X_2 rejects P_i. (This is possible since X_1 and X_2 are epistemically autonomous with respect to each other.) In that case, exactly half of the community accepts P_i while other half rejects P_i. This, I take it, amounts to a maximal amount of disagreement on P_i in the community. And yet, since only two of the community's *n* members are epistemically autonomous, the epistemic

[7] We assume here, for convenience, that *n* is even.

autonomy among members of the group is clearly not maximal (especially when $n >> 3$).

The point of this admittedly contrived example is that, since maximal disagreement does not require maximal autonomy, there is no general connection between degrees of epistemic autonomy and the extent to which a community disagrees of the sort that would have to hold for the 'Millian' solution to account for the value of expert autonomy. There are other, perhaps more realistic, cases that illustrate the same point as well: Suppose, for example, that a group of n agents is divided between two groups of unequal size: (A) a larger subgroup of those who accept P_i, X_1, ..., X_k, and (B) a smaller subgroup of those who reject P_i, X_{k+1}, ..., X_n.[8] Suppose further that all of the P_i-accepting agents in group (A) are fully autonomous while some of the P_i-rejecting agents in group (B) are fully non-autonomous. Now let's consider what happens if one of the fully non-autonomous P_i-rejecting agents in group (B) becomes fully autonomous. Whether such an agent switches from rejecting to accepting P_i will of course depend on how she herself evaluates P_i. Thus it is entirely possible that she will switch – indeed, she is presumably rather likely to do so in our case, given that a majority of those of her peers who evaluated P_i themselves do accept P_i. Now, if this agent does indeed switch to accepting P_i, the majority of P_i-accepters will become larger than it was before, thus *decreasing* the disagreement in the group.

Furthermore, we may suppose that the same process occurs for all the other non-autonomous agents who reject P_i as well. By the same token as before, it is presumably very likely that at least a majority of them switches to accepting P_i, such that the disagreement decreases even further. The upshot, then, is that it is not only possible, but indeed quite likely, that 'autonomizing' agents who were non-autonomous before would significantly *decrease* the community disagreement about P_i. This illustrates once again that there is no general connection between epistemic autonomy and disagreement such that increasing autonomy translates into increased disagreement (or even increased likelihood of increased disagreement). Thus, while it *is* true that epistemic disagreement cannot arise in communities who exhibit no epistemic autonomy at all, the 'Millian' solution to the Puzzle of Epistemic Autonomy fails in so far as it requires there to be a considerably stronger connection between epistemic autonomy and disagreement.

[8] So we are stipulating that $n-(k+1) < k$.

Moving on to the other part of the 'Millian' solution to the Puzzle of Epistemic Autonomy, this claim holds that community disagreement is conducive to reliably obtaining true beliefs. There is a grain of truth to this claim as well. It is certainly *possible* for disagreement to cause or enable a more efficient or more complete discovery of the truth, e.g. through causing agents to gather more evidence than they otherwise would. That much is demonstrated by De Cruz and De Smedt's (2013) case study on *Homo floresiensis*, since 'actually' entails 'possibly'. But in order for the 'Millian' solution to fully account for the value of epistemic autonomy, it is not enough that it is merely *possible* for disagreement to have this effect, since the puzzle is not to account for the fact that epistemic autonomy *can be* epistemically valuable. Rather, if the 'Millian' solution is to really solve the Puzzle of Epistemic Autonomy, it must hold quite generally that disagreements are conducive for the search for truth in the long term.

It is this general claim that I dispute.[9] As I illustrate below, disagreements can stifle the search for truth by focusing intellectual and material resources on attempting to settle disputes that are irresolvable or infertile. In these cases, the search for truth would be better served by exploring an entirely different theoretical alternative that sidesteps or synthesizes the previous alternatives on offer. Specifically, a disagreement prevents such theoretical alternatives from emerging or being seriously considered, due to the relevant community's intellectual resources being too focused on settling the dispute between extant alternatives. So while disagreement on a given issue tends to focus one's cognitive resources on that issue, it is implausible that doing so will generally be conducive to the search for truth.

Let me briefly illustrate with an historical case of disagreement among scientific experts.[10] In the late 19th century, vision studies in Germany were dominated by a wide-ranging dispute between two opposing schools of thought, led by Hermann von Helmholtz and Ewald Hering, respectively. The different theories of human vision endorsed by Helmholtz and Hering have been described as 'empiricist' and 'nativist', respectively, since they differed, *inter alia*, with respect to the extent to which visual perception of space was taken to be acquired (Helmholtz) or inborn (Hering). The two schools disagreed on other less fundamental issues as well, e.g. on whether there are

[9] Note that this is not to deny that there are plenty of examples in which disagreement have brought epistemic benefits (see, e.g., Lougheed 2020: 65-69). Nor is it to deny that disagreement and successful inquiry are statistically correlated (see, e.g. Schulz-Hardt et al 2002, cited in Matheson 2015).
[10] My description of this episode follows Turner 1993, 1994.

three (Helmholtz) or four (Hering) distinct kinds of color receptors in the human eye. Importantly for our purposes, the disagreement between Helmholtz and Hering (and their respective followers) had profound effects on the entire field of vision studies in Germany until well into the 1920s. Considerable resources were spent on trying to convince the proponents of the other side, to little effect.

As historian Steven Turner documents (1993: 90-93), many scientists who were outsiders to the debate felt from the very beginning that the issue had been unhelpfully polarized. Some of these scientists proposed compromises between Helmholtz's and Hering's theories, but these hybrid theories were not seriously considered by proponents of either school of thought, and so did not gain widespread acceptance. In the end, the dispute between Helmholtz's and Hering's theories was not so much resolved as it was *dissolved* when Hering himself and his most enthusiastic followers came to the end of their careers in and around the 1920s. Looking back, Turner (1994: 276-280) suggests that the dispute can be seen as a cautionary tale of the dangers of intense polarization in scientific research. Although it is certainly hard to know what sort of theoretical progress could have been made in this period had the efforts of German vision scientists not been so focused on this particular disagreement, it seems plausible that the Helmholtz-Hering dispute hindered rather than helped the search for truth in the long term.

Of course, this example does not prove any general thesis to the effect that disagreement necessarily stifles the search for truth. What it does is illustrate the point that there does not appear to be any general connection in virtue of which disagreement is conducive to the search for truth. Spending time and resources on finding more evidence to settle some dispute is not always truth-conducive, since the search for truth is sometimes best served by moving on to explore other alternatives (including gathering more evidence relevant to these other alternatives). As I have noted, however, only a general connection between disagreement and truth-conduciveness would support the 'Millian' solution to the Puzzle of Epistemic Autonomy. After all, the solution is meant to apply generally whenever epistemic autonomy is valuable, not just in cases when disagreement contingently happens to be conducive to the search for truth.

I conclude, therefore, that the second causal claim of the 'Millian' solution is as problematic as the first. Hence we must look elsewhere for a solution to the Puzzle of Epistemic Autonomy.

5. Consensus and Altruism

I favor a solution to the Puzzle of Epistemic Autonomy that, in contrast to the two aforementioned solutions, appeals to *altruistic* rather than *egoistic* reasons for exercising epistemic autonomy. The idea, in short, is that critically evaluating an issue will benefit other members of one's epistemic community in virtue of increasing the reliability of consensus positions on that issue. This is epistemically valuable because, and in so far as, these other members of one's community will inevitably have to rely on such consensus positions in their (non-autonomous) assessment of the relevant claim. Thus the idea here is not that you should be epistemically autonomous because and in so far as it – directly or indirectly – brings you some epistemic benefit; rather, you should be epistemically autonomous because and in so far as it benefits other members of your epistemic community.

A key assumption behind this solution is that an epistemic community inevitably exhibits a 'division of cognitive labor' (Kitcher 1990). Some of the community's members will be considered to be more reliable with regard to certain topics, while others will be considered more reliable with regard to other topics. These more reliable individuals are 'experts' with respect to each particular topic. Happily, more often than not, these 'experts' are in fact more reliable than the average person within their domain of alleged expertise. Of course, sometimes it is quite difficult to figure out who are the *most reliable* experts on a particular topic, especially when experts disagree with one another. There are many interesting and difficult questions about what agents should do, epistemically speaking, in situations of that sort.[11] However, it is rarely particularly hard to carve out a broad class of individuals who are almost certainly more reliable than the average person on a particular topic. For example, in the case of academic topics, those who are employed to do research on a particular topic at universities will generally be much more reliable than the average person on that topic. *Mutatis mutandis* for topics that fall under the expertise of other professions.

So with respect to a particular topic – conceived of here as a set of propositions $\{P_i\}$ – an epistemic community can without too much idealization be divided into (a) those that are, and are considered to be, experts on $\{P_i\}$, and (b) those who are not, and

[11] See, e.g., Goldman 2001, Anderson 2011, Guerrero 2017, Dellsén 2018b, and Nguyen 2020.

aren't considered to be, experts on $\{P_i\}$, i.e. those who are 'laypeople' with respect to $\{P_i\}$. In an epistemically well-ordered community, the $\{P_i\}$-laypeople will form their beliefs based on the testimony of the $\{P_i\}$-experts. Note that this is not an elitist or hierarchical system where some people are elevated to the status of epistemic overlords and other people are demoted into an epistemic underclass, since the $\{P_i\}$-experts will inevitably be laypeople with respect to other topics $\{Q_i\}$, on which at least some of the $\{P_i\}$-laypeople will be experts. Indeed, we are all experts with respect to some topics and laypeople with respect to others; no one is simply an expert or a layperson *tout court*.[12]

Now, if you are a layperson with respect to a particular topic $\{P_i\}$ seeking to appeal to the opinions of $\{P_i\}$-experts on that topic, there are roughly three types of situations you might be faced with: (i) Ideally, the $\{P_i\}$-experts will all or overwhelmingly agree on the truth-value of relevant proposition P_i. Consider, for example, the 97% agreement on anthropogenic global warming (AGW), the idea that the recent rise of average surface temperatures on Earth is partially caused by human activity. (This is the type of consensus situation that I take to be most relevant to epistemic autonomy, so I shall return to it shortly.) Now, if the $\{P_i\}$-experts *don't* overwhelmingly agree on the truth-value of P_i, it must be because they either (ii) substantially disagree on P_i, or because (iii) they have yet to come to an opinion on P_i. In either of these latter cases, the opinions of $\{P_i\}$-experts will not be of much use to the layperson, who (unlike the $\{P_i\}$-experts themselves) normally won't be in a position to make fine-grained distinctions in the reliability of different experts. So, again without too much idealization, we can focus on the happier case where the $\{P_i\}$-experts overwhelmingly agree about P_i, i.e. situations of type (i).

Why and when should a $\{P_i\}$-layperson believe the consensus position among $\{P_i\}$-experts in situations of this sort? What makes such a consensus position a reliable guide to the truth? In previous work (Dellsén 2020), I gave a partial answer to this question based on a mathematical fact regarding probabilistic dependence between propositions. The basic idea is this: From the layperson's point of view, experts that are more epistemically autonomous exhibit a greater degree of probabilistic independence in their assessments of a proposition P_i. Given plausible *ceteris paribus* conditions (for details, see Dellsén 2020: 354), this implies that, all other things being

[12] At the very minimum, everyone is an expert with respect to their own lives, personal history, and preferences.

equal, the consensus position of more epistemically autonomous experts will be more likely to be correct than an otherwise identical consensus position held by less autonomous experts. In this sense, epistemic autonomy among experts enhances the extent to which they can be relied upon to deliver correct consensus verdicts to laypeople.

Let me illustrate how this works more concretely by returning to the case of the 97% consensus on anthropogenic global warming (AGW). Why is this consensus position on AGW trustworthy, even from the point of view of someone who has reviewed none of the scientific evidence for AGW and is in no position to evaluate the theory on the basis of that evidence? Well, suppose – counterfactually! – that AGW was initially proposed by a small group of powerful scientists whose influence on the various subfields of climate science was so immense that all other climate scientists accepted AGW entirely on the basis of their say-so. In this counterfactual scenario, the 97% consensus on AGW would provide laypeople with no more reason to believe it than if only the original group had announced its acceptance of the theory. After all, in such a scenario, the other thousands of climate scientists would have accepted AGW even if the original group had made a mistake in their evaluation of the theory (or was part of some nefarious conspiracy) and the actual scientific evidence for it was weak or non-existent. So in this type of case, the additional thousands of scientists who adopt the same position on AGW as the original group would provide no additional reason to believe AGW whatsoever.

Fortunately, this is not the actual situation with regard to AGW. There is no group of purportedly-infallible climate scientists whose say-so is automatically and uncritically channeled by other climate scientists. Rather, the climate science community is much closer to one in which each scientist critically evaluates AGW for themselves and reaches a conclusion (sometimes on the basis of the same evidence as their peers, but more often on the basis of *different* evidence and even different *types* of evidence). Of course, no actual epistemic community is such that *every single one* of its members is *completely* autonomous with respect to any proposition. But a large reason why we ought to trust the climate science community's evaluation of AGW is that its thousands of members are sufficiently epistemically autonomous for it to be nearly impossible that the community would reach a consensus position on AGW unless it was true.

So, to summarize, epistemic autonomy is valuable because, and in so far as, it leads experts on a particular topic to be reliable in their consensus positions, which is in turn valuable because, and in so far as, laypeople with respect to the topic rely on such expert consensus positions to form reliably true beliefs. Now, this is by no means an unqualified defense of epistemic autonomy for all agents in all circumstances, for it does not apply to agents in so far as they consider themselves to be laypeople rather than experts with regard to a particular set of propositions $\{P_i\}$. (Indeed, this solution *presupposes* that at least some of these $\{P_i\}$-laypeople *don't* exercise epistemic autonomy with regard to $\{P_i\}$ – since if they would, there would be no point to having a reliable expert consensus on $\{P_i\}$.) Rather, the solution to the Puzzle of Epistemic Autonomy that I have sketched here applies specifically to those who serve as 'experts' in our community: If you are someone who is consulted as an expert on $\{P_i\}$, you should exercise epistemic autonomy with respect to $\{P_i\}$ since that makes it more reliable for other people to rely (non-autonomous) on the consensus position reached by you and your fellow $\{P_i\}$-experts.

In my view, this is a plausible and desirable restriction to epistemic autonomy. A solution to the Puzzle of Epistemic Autonomy ought not entail that all agents should be epistemically autonomous with respect to all topics. Such a solution would 'prove too much'. Instead, a sensible solution should distinguish between the epistemically desirable and undesirable instances of epistemic autonomy in a plausible way. The current solution to the Puzzle of Epistemic Autonomy does precisely that. For instance, it implies that those of us who are laypeople with respect to AGW should not base our beliefs on our own autonomous assessment of the theory; rather, we should defer to the experts, i.e. climate scientists. On the other hand, to the extent that we are considered experts on a given claim or topic – e.g., as the contributors to this volume are taken to be experts on epistemic autonomy itself – the current solution implies that one should resist deferring to other experts on that topic and instead make up one's own mind. And the *reason* one should do this is the altruistic one that making up one's own mind benefits one's fellow epistemic agents. Believe me, I'm an expert – or don't, if you are one.[13]

[13] Many thanks to Zach Barnett, Kirk Lougheed, Federica Malfatti, and Jon Matheson for very helpful feedback on drafts.

References

Anderson, E. (2011). Democracy, Public Policy, and Lay Assessments of Scientific Testimony. *Episteme* 8: 144-164.

Bird, A. (2007). What Is Scientific Progress? *Nous* 41: 64–89.

Bourget, D. (2017). The role of consciousness in grasping and understanding. *Philosophy and Phenomenological Research*, 95: 285-318.

Boyd, K. (2017). Testifying Understanding. *Episteme* 14:103–127.

Cavender, N. and Kahane, H. (2009). *Logic and Contemporary Rhetoric: The Use of Reason in Everyday Life, Eleventh Edition*. Wadsworth, Belmont.

Constantin, J., Grundmann, T. (2018). Epistemic authority: Preemption through source sensitive defeat. *Synthese*. DOI: 10.1007/s11229-018-01923-x.

De Cruz, H. and De Smedt, J. (2013). The value of epistemic disagreement in scientific practice: The case of homo floresiensis. *Studies in History and Philosophy of Science*, 44:169–177.

de Regt, H. W. (2017). *Understanding Scientific Understanding*. Oxford University Press, Oxford.

Dellsén. F. (2016). Scientific Progress: Knowledge versus Understanding. *Studies in History and Philosophy of Science* 56: 72-83.

Dellsén, F. (2017). Understanding Without Justification or Belief. *Ratio*, 30:239–254.

Dellsén, F. (2018a). Deductive Cogency, Understanding, and Acceptance. *Synthese*, 195:3121–3141.

Dellsén, F. (2018b). When Expert Disagreement Supports the Consensus. *Australasian Journal of Philosophy*, 96:142-156.

Dellsén, F. (2020) The Epistemic Value of Expert Autonomy. *Philosophy and Phenomenological Research*, 100:344–361.

Dellsén, F. (forthcoming). Beyond Explanation: Understanding as Dependency Modeling. *The British Journal for the Philosophy of Science*. DOI: 10.1093/bjps/axy058.

Descartes, R. (1985/1628). *Rules for the Direction of the Mind*. In Cottingham, J., Stoothoff, R., and Murdoch, D., editors, The Philosophical Writings of Descartes, volume I. Cambridge University Press, Cambridge.

Elgin, C. Z. (2010). Persistent Disagreement. In R. Feldman and T. Warfield (eds.), *Disagreement*. Oxford: Oxford University Press, pp. 53-68.

Elgin, C. Z. (2017). *True Enough*. MIT Press, Cambridge, MA.

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

Grimm, S. (2006). Is Understanding a Species of Knowledge? *British Journal for the Philosophy of Science*, 57:515–535.

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

Guerrero, A. (2017). Living with Ignorance in a World of Experts. In R. Peels (ed.), *Perspectives on Ignorance from Moral and Social Philosophy*. New York: Routledge, pp. 168-197.

Hills, A. (2009). Moral testimony and moral epistemology. *Ethics*, 120:94–127.

Hills, A. (2016). Understanding Why. *Nous*, 50:661–688.

Huemer, M. (2005). Is Critical Thinking Epistemically Responsible? *Metaphilosophy*, 36:522-531.

Kant, I. (1991/1784). An Answer to the Question: What is Enlightenment? In Reiss, H., editor, *Political Writings, 2nd edition*. Cambridge: Cambridge University Press, pp. 54–60.

Khalifa, K. 2013. 'Understanding, Grasping and Luck.' Episteme, 10: 1–17.

Khalifa, K. 2017. *Understanding, Explanation, and Scientific Knowledge*. Cambridge: Cambridge University Press.

Kitcher, P. (1990). The Division of Cognitive Labor. *Journal of Philosophy*, 87:5–21.

Locke, J. (1975/1689). *An Essay Concerning Human Understanding*. Clarendon Press, Oxford.

Lougheed, K. 2020. *The Epistemic Benefits of Disagreement*. Cham: Springer.

Malfatti, F. I. (2019). On Understanding and Testimony. *Erkenntnis*. DOI: 10.1007/s10670-019-00157-8.

Malfatti, F. I. (2020). Can Testimony Transmit Understanding? *Theoria*. DOI: 10.1111/theo.12220.

Matheson, J. (2015). Disagreement and the Ethics of Belief. In Collier, J. (ed.), *The Future of Social Epistemology: A Collective Vision*. Lanham: Rowman and Littlefield, pp. 139-148.

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

Nguyen, T.C. (2020). Cognitive islands and runaway echo chambers: problems for epistemic dependence on experts. *Synthese* 197: 2803-2821.

Nussbaum, M. (2017). *Not for Profit: Why Democracy Needs the Humanities (Updated Edition)*. Princeton University Press, Princeton.

Potochnik, A. (2017). *Idealization and the Aims of Science*. University of Chicago Press, Chicago, IL.

Pritchard, D. (2009). Knowledge, Understanding, and Epistemic Value. In O'Hear, A., editor, *Epistemology (Royal Institute of Philosophy Lectures)*, pages 19–43. Cambridge University Press, Cambridge.

Pritchard, D. (2010). Knowledge and Understanding. In *The Nature and Value of Knowledge: Three Investigations*, pages 3–90. Oxford University Press, Oxford.

Schulz-Hardt, S., M. Jochims, and D. Frey. (2002). Productive Conflict in Group Decision Making: Genuine and Contrived Dissent as Strategies to Counteract Biased Information Seeking. *Organized Behavior and Human Decision Processes* 88: 175-195.

Strevens, M. (2013). No Understanding Without Explanation. *Studies in History and Philosophy of Science*, 44:510–515.

Turner, R. S. (1993). Vision Studies in Germany: Helmholtz versus Hering. *Osiris* 8:80–103.

Turner, R. S. (1994). *In the Eye's Mind: Vision and the Helmholtz-Hering Controversy*. Princeton University Press, Princeton.

Wilkenfeld, D. A. (2013). Understanding as Representation Manipulability. Synthese, 190:997–1016.

Zagzebski, L. (2007). Ethical and Epistemic Egoism and the Ideal of Autonomy. *Episteme*, 4:252–263.

Zagzebski, L. (2008). *On Epistemology*. Belmont, CA: Wadsworth.