Epistemic Paternalism

Kristoffer Ahlstrom-Vij

University of Kent, Canterbury

hka@kent.ac.uk

Word count: 7,562

1. The Case for External Constraints

It's a well-established fact that we often reason by way of *heuristics*, or subconscious rules of thumb, rather than through the systematic application of formal rules or principles of logic, statistics, probability theory, and so forth (Kahneman 2011; Gilovich et al. 2002; Kahneman et al. 1982). Such heuristics typically operate by making certain *assumptions*, for example to the effect that all information necessary for making a sound judgment is readily available or otherwise vividly presented to the agent (*availability heuristic*), or that any given sample, even if small, will be representative of the population from which it is drawn (*representativeness heuristic*). In many cases, these assumptions don't present any problem. Indeed, on certain tasks heuristical reasoning outperforms more labor-intensive reasoning strategies by a comfortable margin (Gigerenzer et al. 1999), which might in some cases be explained with reference to the adaptive nature of the relevant assumptions (Cosmides and Tooby 1996).

At the same time, the question remains: adapted to *what*? The time span separating the pre-historic from the modern world is too short for any evolutionary pressure to have brought our cognitive apparatus up to speed with a wide variety of modern challenges. Consequently, even if largely adaptive, we have good reason to worry about heuristical reasoning in many contexts making for *bias*, or systematic reasoning mistakes arising when the assumptions that the relevant heuristics are operating on simply don't hold.

Of course, any concern about bias would be greatly diminished were it simply a matter of being more careful and vigilant in our thinking. One way to flesh out this thought is in terms of what we may refer to as *the self-correction strategy*, on which the individual agent corrects for bias on her own accord. However, there are two problems for this strategy. The first problem is one of *motivation*, arising out of the fact that any attempt to deal with bias has to take into account not only that we are biased, but also that we suffer from what Emily Pronin and colleagues (2002) have referred to as a 'bias blind spot', on account of which we tend to underestimate the extent to which we are prone to bias. This blind spot should be understood in the context of the well-known psychological fact that, depressed people aside (Taylor and Brown 1988), we tend to rate ourselves as above average on desirable traits (Alicke 1985; Brown 1986). This overconfidence extends to our evaluations of our own epistemic capabilities. As Pronin (2007) notes in an overview, "people tend to recognize (and even overestimate) the operation of bias in human judgment—except when that bias is their own" (37).

Consequently, the first problem facing the self-correction strategy is that, whether or not there are corrective measures available, each and every one of us will tend not to see the point of taking corrective measures, on account of our bias blind spots. But let's assume that there's a way around this problem. Still, merely being *motivated* to correct for bias is not enough—additionally, we need to do so *successfully*. This is the second problem for the self-correction strategy. Solving that problem requires doing two things. First, we need to correct for bias when and only when we are in fact biased. This poses a challenge of *bias identification*. The challenge is that the most obvious way to look for bias is by introspecting, while the bulk of the operations that would require scrutiny simply aren't introspectively accessible (Wilson 2002). And, as pointed out by Wilson and Brekke (1994), while we often have access to the outputs of those operations, bad judgments, unlike bad food, do not smell. But say we find a way to overcome the challenge of bias identification. We now need to correct to and only to

the extent needed to remove any bias. This is the challenge of *proper correction*. Such correction is difficult on account of the risks of engaging in *insufficient* correction or *over*correction (Petty and Wegener 1993).

Of course, none of this suggests that it is *impossible* for people to meet the challenges of bias identification and proper correction. Still, it seems clear that, even if we assume that the relevant agents are at all motivated to engage in bias correction—which is far from a trivial assumption, as we've seen—there are substantial challenges they need to meet when it comes to doing so successfully. So what's a more promising strategy?

The strategy to be pursued in what follows focuses not on the correction but on the *prevention* of bias. This is in line with the de-biasing literature generally. For example, Wilson and colleagues (2002) note that "[t]he best way to avoid biased judgments and emotions is exposure control" (195). In light of the problem of motivation outlined earlier, the most promising way to achieve such control involves imposing *external* constraints on the agent to shield her from biasing information. For example, consider the practice on the part of US judges to withhold certain types of information from the jurors, such as character evidence or evidence about past crimes, on the assumption that the jurors are likely to systematically overestimate the probative value of such information. Consequently, on the *US Federal Rules of Evidence*, the mere fact that a piece of evidence is *relevant*, in making the hypothesis about guilt more or less likely than it otherwise would have been, is not a sufficient condition for presenting it to a jury. It also matters whether jurors are likely to *gauge* that relevance properly. If not, the presiding judge may withhold the information. This type of evidence control serves to illustrate one particular type of external constraint, namely an external constraint on *information access*, restricting the choices the agent can make when it comes to what information to bring to bear on whatever matter she happens to be considering.

2. Defining Epistemic Paternalism

The previous section offered some reasons for imposing external constraints on information access as a form of bias exposure control, in light of challenges for the idea of having individual agents correct for bias on their own. In this section, I'll make the case that such external constraints are properly referred to as *epistemically paternalistic*. To make that case, we need to get clearer on what epistemic paternalism is. I'll suggest that there are three jointly sufficient conditions: *the interference condition*, *the non-consultation condition*, and *the improvement condition*. Let's consider each in turn.

One central feature of a paternalistic practice is that it constitutes an interference with the actions of another for her own good. In cases of epistemic paternalism, what's interfered with is someone's *inquiry*, understood here as involving the accessing, collection and evaluation of information. I will take it that someone is *interfering* with the inquiry of another if the former is compromising the latter's freedom to conduct inquiry in whatever way she sees fit, and that someone is *free* to conduct inquiry thus if there are no constraints imposed by others on her ability to access, collect and evaluate information. When a practice interferes with someone's inquiry by compromising her freedom in this manner that practice satisfies *the interference condition* on epistemic paternalism.

When judges withhold certain kinds of evidence from the jurors, the jurors' freedom to access whatever information they deem significant in a manner free of constraints imposed by others is compromised. As such, the relevant practice satisfies *the interference condition*. In light of this, it should come as no surprise that this practice has been termed paternalistic in the literature (e.g., Laudan 2006) and, in one instance, *epistemically* paternalistic (Goldman 1991). Whether this practice is in some relevant sense *justified* is, of course, another matter, and one that we'll find reason to return to in Section 5. For now, let's move on to the second condition on epistemic paternalism.

There are reasons why people have found paternalism problematic. For one thing, there is something *arrogant* about paternalistic interference. However, it's not necessary that the person interfered with *objects* to the interference for it to qualify as paternalistic (Dworkin 2010). Consider an example from Shiffrin (2000): if I believe a friend of mine to be financially irresponsible, I might intercept a credit card offer that he gets in the mail. I am not thereby doing something that he objects to, since he is aware neither of the offer nor of my intercepting it. Still, it seems right to say that I am acting paternalistically. This suggests that what makes a practice paternalistic is not that those interfered with are objecting, but that they are not *consulted*. It might be objected that paternalistic interference requires that those interfered with *would* object, had they been consulted. So, assume that my friend actually would have *agreed* that my interference was called for, had I bothered to consult him. Since I don't, it seems that I am still acting paternalistically. As such, the idea that what matters for epistemically paternalistic interference is whether or not those interfered with have been consulted stands. Hence, *the non-consultation condition*.

Do judges withholding biasing evidence from jurors satisfy the non-consultation condition? At no point have they consulted those from whom they are withholding information as to whether the information should be withheld. Consequently, the relevant practice satisfies the non-consultation condition.

We started out by saying that a paternalistic practice involves an interference with the doings of another for her own good. In the case of epistemic paternalism, the relevant good is an *epistemic* good, in turn a function of either succeeding or standing a good chance of succeeding in forming true belief and avoiding false belief (Ahlstrom-Vij 2013*a*). That said, there are a number of epistemically relevant dimensions along which an agent may be better or worse off in the relevant sense. I will focus on two such dimensions: reliability and (question-answering) power (Goldman 1992). Reliability is a matter of avoiding error by generating a high *ratio*

of true to false belief. Power is the ability to form a large *number* of true beliefs that constitute correct answers to whatever questions are facing the agent.

How are we to weigh improvements in reliability against improvements in power? I will not attempt to answer this question here. Instead, borrowing a term used in contexts of interpersonal comparisons of welfare, I will talk in terms of (intra-personal) *epistemic Pareto improvements*, that is, improvements along one epistemic dimension that do not entail a deterioration with respect to any other epistemic dimension. In what follows, I will focus on Pareto improvements in *reliability*, the simple reason being that this is a kind of improvement about which we have relevant empirical data. That, moreover, seems a reasonable description of what's going on in cases of evidence control: by withholding biasing information, judges are attempting to increase jurors' reliability, without thereby making them worse off along some other epistemically relevant dimension. The practice thereby satisfies *the improvement condition*.

Does it need to be the case that the relevant agent is interfered with *solely* for the purpose of making her epistemically better off for the practice involved to qualify as epistemically paternalistic? No, and epistemic paternalism is in that respect a *mixed* form of paternalism (Feinberg 1986). By interfering with someone's inquiry for the purpose of making them epistemically better off, we might very well also be looking to make other people better off in *non*-epistemic terms. For example, when withholding information from jurors, judges are motivated not solely by a desire to make the jurors epistemically better off, but also by a desire to protect the defendant's welfare while doing right by those wronged.

However, this might be taken to raise an objection: aren't we in that case really interfering, *not* for the good of those interfered with, but *rather* for the good of others? Differently put, isn't the *real* reason for interference the good of others? If it is, epistemic paternalism arguably isn't a form of paternalism at all. That, however, seems incorrect. Something is a real reason

for someone doing something when it's part of a plausible explanation of *why* that someone did what she did. In the case of epistemically paternalistic practices, the fact that those interfered with are made epistemically better off forms part of the explanation of why we interfere in the manner we do. Part of the reason judges interfere with the information available to the jurors is that it makes them epistemically better off. The judges' *ultimate* aim might not be the jury's epistemic improvement, but rather pertain to the defendants and those wronged. Still, the fact remains that the judges are interfering *both* for the (epistemic) good of those interfered with *and* for the (non-epistemic) good of others, on the grounds that achieving an improvement in the former is an *instrument* to securing the latter.

However, someone might maintain that paternalistic interference needs to be ultimately concerned with the good of those interfered with. But that's implausible. Another example from Shiffrin helps us see why:

> Suppose a park ranger has the power to refuse permission to climb a steep, dangerous mountain path. If the ranger refuses to allow a person to climb simply because the (fully informed, competent) person might hurt himself and that would be bad for him, that refusal would be paternalist, I think. Suppose the ranger says, "Of course, you may take whatever risks you want with your life, but I refuse permission because you might die and leave your spouse grief-stricken." Such a refusal also seems paternalist. The ranger is substituting his judgment about how the climber should treat her spouse and conduct her marriage. The ranger is taking over the climber's marriage, a bit, on the implicit grounds that his moral judgment of how to conduct the relation is correct and the climber's is incorrect (Shiffrin 2000: 217).

In the second case described by Shiffrin, the ultimate aim of the ranger's interference is the welfare of the climber's spouse; his concern for the welfare of the climber is instrumental to this aim. But, according to Shiffrin, the interference is still paternalistic since "it involves a person's aiming to take over or control what is properly within the agent's own legitimate domain of judgment or action" (216). Something similar is going on in the case of evidence control. The jurors have a legitimate power to act as the triers of fact. Through her interference, the judge infringes on that power. That, on Shiffrin's account, is what makes her interference paternalistic. But on my account, that's not the whole story. Unlike Shiffrin, I don't find plausible the claim that "the paternalist may be *solely* and directly concerned with the third party's welfare" (216; my emphasis). Say that the judge, worried about the epistemic caliber of the jury not being high enough to safeguard the defendant and do right by those wronged, dismisses the jury and takes it upon herself to act as the sole trier of fact. In so doing, she certainly takes over the control of the jury's legitimate domain of judgment and action—but she doesn't, I submit, act paternalistically. What's missing is a concern, even an instrumental one, for the epistemic good of the jurors. That concern is present in the original case of evidence control: the judge interferes within the legitimate domain of the jury, but does so with an eye towards making the jury epistemically better off.[1]

So, to sum up, in the case of epistemically paternalistic practices, it's *not* the case that we are interfering for the good of others rather than for the good of those interfered with. We are interfering for the good of both. More specifically, we are interfering on the grounds that making those interfered with (epistemically) better off is an instrument towards making others better off. And, as we have just seen in relation to Shiffrin's account of paternalism, the fact that our *ultimate* aim in so interfering thereby isn't to make those interfered with better off isn't sufficient to show that the relevant form of interference isn't paternalistic.

3. Epistemic Paternalism and Autonomy

The previous section suggested that we practice epistemic paternalism when interfering with the inquiry of another for her own epistemic good without consulting her on the issue. On this definition, the current practice of evidence control qualifies as a form of epistemic paternalism. It doesn't follow from any of this, of course, that we *should* practice epistemic paternalism. In the present section, we'll consider the strongest form of argument suggesting that we should *not*, on account of considerations about personal autonomy.

The most famous objection of this kind was offered by Joel Feinberg (1986). As Feinberg saw it, "[t]he anti-paternalist [...] must not only argue against particular legislation with apparently paternalistic rationales; he must argue that paternalistic reasons never have any weight on the scales at all" (25). This is so because "they are morally illegitimate or invalid reasons by their very natures, since they conflict head on with defensible conceptions of personal autonomy" (26). At the heart of that conception is the idea that

> [...] *respect for a person's autonomy is respect for his unfettered voluntary choice as the sole rightful determinant of his actions except where the interest of others need protection from him*. Whenever a person is compelled to act or not to act on the grounds that he must be protected from his own bad judgment even though no one else is endangered, then his autonomy is infringed (Feinberg 1986: 68).

In other words, someone's autonomy is infringed when the *sole* reason for interference is that it is in his own interest, and it thereby is not the case that the interests of others are being factored in. But notice that this does not imply that paternalistic reasons—that is, reasons in terms of the good of those interfered with—are always invalid. At best, Feinberg's notion of

autonomy serves to rule out a certain kind of paternalistic *practice*, namely one motivated *exclusively* with reference to reasons pertaining to the good of those interfered with. As we saw earlier, however, the kind of epistemically paternalistic practices that concern us here are not of this kind. In so far as they are justified, they are justified with reference to a concern *both* for the epistemic good of those interfered with, and for the non-epistemic good of others. That's why Feinberg's notion of autonomy is compatible with epistemic paternalism.

Notice, however, that Feinberg merely provides a *sufficient* condition on autonomy infringement. Maybe there are other ways to violate people's autonomy, not captured by Feinberg's account. On this point, consider Joseph Raz's (1986) notion of autonomy, on which '[o]ne is autonomous if one determines the course of one's life by oneself' (407). A respect for autonomy rules out coercion or manipulation, both of which constitute ways for one person to impose her will on others (378). In the case of coercion, this is done through the reduction or removal of a person's options. Manipulation, by contrast, "perverts the way that [the] person reaches decisions, forms preferences, or adopts goals" (377-8).

If a mere manipulation of available options—let alone the removal of such options through coercive means—constitutes a violation of autonomy, it seems that paternalism necessarily violates people's autonomy. Or does it? Not according to Raz:

> [P]aternalism affecting matters which are regarded by all as of merely instrumental value does not interfere with autonomy if its effect is to improve safety, thus making the activities affected more likely to realize their aim. There is a difference between risky sports, e.g., where the risk is part of the point of the activity or an inevitable by-product of its point and purpose, and the use of unsafe common consumer goods. Participation in sporting activities is intrinsically valuable. Consumer goods are normally used for instrumental reasons (Raz 1986: 422-3).

As argued above, the epistemic goods promoted through epistemic paternalism are instrumental rather than ultimate or intrinsic goods. In this respect, the practices subject to epistemic paternalism have more in common with the use of consumer goods than with risky sports. For example, unlike in the case of risky sports, it is no essential component of legal proceedings that things might go wrong, and the means involved sometimes fail to secure the ultimate goods. That's why epistemic paternalism also is compatible with Raz's notion of autonomy.

4. Justifying Epistemic Paternalism

In order to provide a *defense* of epistemic paternalism, it's not sufficient to show that epistemic paternalism isn't inherently objectionable. It also needs to be shown that there are situations in which we are justified in practicing epistemic paternalism. For that purpose, I'll be offering two jointly sufficient conditions for justified epistemically paternalistic interference. *The alignment condition* pertains to the interplay between our reasons and helps us avoid some issues arising when we try to weigh different kinds of reasons against each other. *The burden-of-proof condition* speaks to the circumstances under which one's beliefs about the desired effects are justified. We'll consider these conditions in turn.

Let us return once more to the case of evidence control. The judge withholds biasing information from the jurors in order to make them epistemically better off, because doing so serves the non-epistemic end of protecting the welfare of the defendant and doing right by those wronged. In other words, the paternalistic measure in question involves two serially ordered motivations, picking out means and ends, respectively. But we can certainly imagine cases where epistemic and non-epistemic motivations do not line up so nicely. Consider, for example, a society exercising total control over the minutest details of the epistemic undertakings of their citizens for purposes of making them better off along epistemic dimensions. Even if successful,

we might feel that such a regime would be far too intrusive, and that we on that account would have good reason to reject it. Here, we could talk in terms of different kinds of reasons being *weighed* against each other, and say that paternalistic interventions are justified only if the reasons for intervening outweigh those against. This is the strategy employed by Peter de Marneffe (2010) in his discussion of paternalistic restrictions on prostitution. This is not the strategy I'll be employing here, however, the reason being that, while it's fairly easy to make sense of what is for reasons to have *valence*—being for or against things—it's often more difficult to assign *weights* specific enough to help determine what outweighs what.

What's an alternative strategy, then? The one I'll be pursuing can be framed in terms of the following condition:

> **The Alignment Condition**: The epistemic reasons we have for instituting the relevant epistemically paternalistic practice, on the grounds that it will lead to an epistemic improvement, are *aligned* with our non-epistemic reasons on the issue. Two or more reasons are aligned if and only if they are (*a*) reasons for the same thing, or, failing that, (*b*) silent on the issue, by not constituting reasons either way on the matter.

One benefit of the alignment condition is that it only requires that reasons have valence. The relative weights of reasons do not need to be factored in to determine whether the condition is satisfied. Note, however, that the alignment condition does not provide a *necessary* condition on justified epistemic paternalism. We can imagine epistemically paternalistic practices that fail to satisfy the alignment condition, but nevertheless are justified on weighing grounds. The alignment condition also doesn't offer a *sufficient* condition on epistemically paternalistic interventions. While it guarantees a certain *harmony* among reasons, it doesn't entail that we are

justified in believing that the relevant form of interference will actually have the intended effects. As it happens, this gets to a major worry about paternalistic interference, which will motivate our second condition.

The relevant worry was voiced already by John Stuart Mill, who suggested that "the strongest of all arguments" against paternalistic interference is that, if we attempt to interfere, the odds are that we will do so "wrongly and in the wrong place":

> On the question of social morality, of duty to others, the opinion of the public, that is, of an over-ruling majority, though often wrong, is likely to be still oftener right; because on such questions they are only required to judge of their own interests; of the manner in which some mode of conduct, if allowed to be practiced, would affect themselves. But the opinion of a similar majority, imposed as a law on the minority, on questions of self-regarding conduct, is quite as likely to be wrong as right; for in these cases public opinion means, at the best, some people's opinion of what is good or bad for other people; while very often it does not even mean that; the public, with the most perfect indifference, passing over the pleasure or convenience of those whose conduct they censure, and considering only their own preference (Mill 1989/1859: 83-4).

There are two reasons that this argument does not apply in the case of epistemic paternalism. First, when it comes to epistemic goods, it's *not* the case that each person necessarily knows her own good best. (Young 2008, de Marneffe 2006, and Hart 1963 offer some reasons to think the same goes in non-epistemic cases.) Indeed, given our introspective limitations and tendencies for overconfidence, it cannot be ruled out that the individual agent might in many cases be the *worst* judge on the matter of whether she would undergo an epistemic Pareto improvement on account of some intervention.

Second, the grounds on which it should be judged whether someone is likely to be made better off epistemically through some form of interference are not majority votes, but our best empirical evidence on the issue. More specifically, consider the following:

> **The Burden-of-Proof Condition**: A case can be made that available evidence indicates that it is highly likely that everyone interfered with in the relevant manner is or will be made epistemically better off for being interfered with thus, compared to relevant alternative practices.

What constitutes relevant alternative practices will vary from case to case, but falls into two broad categories. In cases where we are considering implementing a new paternalistic practice to replace the prevailing one, the relevant alternative is the prevailing practice. In cases where we are attempting to justify a practice already in place, the relevant alternatives are whatever practices figure as prominent alternatives, paternalistic or otherwise.

A word is in order on the claim that the evidence needs to indicate that *everyone* interfered with will be made epistemically better off, as it might be taken to raise an objection. Consider a case wherein the burden-of-proof condition is satisfied, and a case consequently can be made that it is highly likely that everyone interfered with will be made epistemically better off for being interfered with. This is compatible with some of those interfered with actually being made worse off. So, let us assume that a small minority actually is. Doesn't this provide an epistemic reason *against* interference, and imply that the alignment condition thereby is not satisfied? And, if it does, can we *ever* assume the alignment condition to hold, given that we typically cannot rule out that *some* people might be made worse off by being interfered with? This is *the problem of the epistemic outlier*.

Let us spell out the imagined scenario in more detail. The scenario is one where the practice satisfies the burden-of-proof condition, while the fact that a minority of those interfered with—the epistemic outliers—will be made worse off is unknown to us. If it were not, the burden-of-proof condition would no longer be satisfied. To determine if the adversely affected minority nevertheless provides us with a reason *not* to interfere, we need to consider on what grounds they are adversely affected. One possibility is that the adverse effect is merely *accidental*. Accidental effects are low probability effects. When it comes to the kind of *ex ante* justifications at issue here, however, we have reason to do whatever is highly likely to generate a good epistemic outcome on our evidence. In the case at hand, our evidence suggests that there is a high likelihood that each person benefits epistemically. Consequently, the mere fact that someone will be affected in unintended and purely accidental ways does *not* provide an epistemic reason against interfering.

Another possibility is that the adverse effect is *not* accidental. Perhaps the epistemic outliers possess some superior epistemic capability that the interference prevents them from relying on. But here, too, it is not clear that this gives us any epistemic reason not to interfere. It is certainly *possible* that there are people that will be systematically disadvantaged by being interfered with in epistemically paternalistic ways. But keep in mind the perspective from which *ex ante* justifications are provided, namely one from which we in effect are placing an empirical bet on what will have the best effect for those interfered with. We could place our bet on the basis of a mere possibility. Or we could acknowledge that it is highly unlikely that people will be disadvantaged, given what we know about our tendencies for bias and overconfidence, as well as the resulting benefits of external constraints, and instead place our bet on the basis of the available evidence. If the burden-of-proof condition is satisfied, that evidence suggests that it is highly likely that the relevant interference will have the intended effect.

Both of these responses assume that we do not *know* that some people are or will be adversely affected. Would finding that out defeat our justification for interfering? Yes, it would, but not on account of the scenario failing to satisfy the alignment condition. As previously noted, if we were to find out that not everyone does or will benefit from the interference, the burden-of-proof condition would no longer hold. But in that case, the response is not necessarily to back away from interference, but to adjust its scope in such a way that those we know to be adversely affected no longer are interfered with in the relevant way.

Hence, the alignment and burden-of-proof are and remain plausible conditions. Of course, no conclusive case has been made to the effect that their combination provides a sufficient condition for justified epistemic paternalism. In the next and final section, we consider whether the epistemically paternalistic practice of evidence control satisfies the conditions. If it does, and it's not obviously unjustified on intuitive grounds, this not only provides evidence for the joint sufficiency of the conditions, but also for the main thesis of this chapter, to the effect that we are sometimes justified in interfering with the inquiry of another without her consent but for her own epistemic good.

5. Epistemic Paternalism Defended

The question for the present section is whether the type of evidence control discussed above is a *justified* epistemically paternalistic practice. To put the question in terms of the conditions outlined in the previous section: (*a*) can a case be made that available evidence indicates that it is highly likely that everyone interfered with in the relevant manner is made epistemically better off for being interfered with thus, compared to relevant alternative practices; and (*b*) are the epistemic reasons we have for instituting the relevant practice aligned with our non-epistemic reasons on the issue?

Let us start by considering what the relevant alternatives to prevailing practices would be. According to Larry Laudan (2006), one of the most forceful critics of evidence control, the alternative would be a practice requiring that 'the only factor that should determine the admissibility or inadmissibility of a bit of evidence is its relevance to the hypothesis that a crime occurred and that the defendant committed it' (25). Moreover, Laudan's motivation for this claim is epistemic: 'Paternalistically coddling jurors by shielding them from evidence that some judge intuits to be beyond their powers to reason about coherently is not a promising recipe for finding out the truth' (23).

I think Laudan is mistaken on this point, at least as it pertains to character evidence. To see why, we need to consider the fact that there is a psychological asymmetry in how jurors treat character evidence that favors *negative* character evidence framed in terms of *particular* actions over other kinds of character evidence (Maeder and Hunt 2011). These results are in line with studies suggesting both that anecdotal information in terms of *specific* acts tend to be considered more diagnostic than base rate information (Borgida and Nisbett 1977), and that *immoral* behavior is taken to be more diagnostic of an individual's character than is moral behavior (Lupfer et al. 2000).

Moreover, this psychological asymmetry is a symptom of people *overestimating* the probative value of particular negative character evidence. To see why, consider that what a juror swayed by particular character evidence does is reason (often unconsciously) from evidence about past actions to what social psychologists would call a *personality trait*, consisting in a general disposition to behave in a certain manner. Then, the juror factors in the nature of that trait in her decision about the guilt of the defendant. The problem is that such traits are not particularly predictive of how people behave across different situations, on account of not amounting to what John Doris (2002) calls *robust* traits. The characters we instantiate simply do not manifest a sufficiently high degree of cross-situational consistency. The extent to which

we are generous, mean, helpful or what have you, on any given occasion, owes more to surprisingly small and seemingly irrelevant differences in situation than to anything like an inherent character.

This is relevant to the epistemic situation of the juror. In cases of negative character evidence framed in terms of particular acts, we seem to invest a significant amount of credence, despite such evidence not providing particularly valuable information about other actions. In that sense, we invest too much credence into negative character evidence, if framed in terms of particular actions. This is sufficient to show that Laudan's (2006) claim—that 'the only factor that should determine the admissibility or inadmissibility of a bit of evidence is its relevance to the hypothesis that a crime occurred and that the defendant committed it' (25)—is not epistemically well motivated. Introducing character evidence means introducing a kind of evidence that will make it harder for jurors to evaluate the objective weight of the evidence properly and arrive at an informed verdict. That is why relevance is *not* sufficient for admissibility, even on purely epistemic grounds.

In light of current evidence, it thereby seems a case can be made that it is highly likely that jurors will be made more reliable in so far as there are some restrictions on the admissibility of character evidence, as per *the Federal Rules of Evidence*, compared to there being no such restrictions. But are they thereby made worse off along some other epistemic dimension, such as power, by forming fewer true beliefs on the question of guilt?

In one sense, there are only two beliefs available to the jurors: that the defendant is guilty or that the defendant is not guilty. And if so, admitting or not admitting character evidence will make no difference to the number of beliefs formed by the jury. But it might be argued that there are many different kinds of beliefs that jurors might form in relation to the question of guilt, including beliefs about the trustworthiness of witnesses, the probative value of the evidence introduced, and so on. Introducing less evidence might certainly have the jury form fewer

such beliefs, for the simple reason that introducing less evidence means that there are fewer things to form beliefs about. Still, if the previous arguments for the idea that introducing character evidence makes jurors less reliable are correct, it is not clear that putting restrictions on the admissibility of such evidence will make them form fewer *true* beliefs.

Consequently, the burden-of-proof condition is met. But what about the alignment condition? To see why our epistemic and *moral* reasons are aligned, remember that, when a judge withholds certain evidence on the grounds that it might bias their judgment, she does this to safeguard the welfare of the defendant while doing right by those wronged. This makes for straightforward alignment: Safeguarding the welfare of the defendant and doing right by those wronged requires convicting all and only those who are in fact guilty, and the best way of approximating this ideal is to ensure that those making judgments about guilt evaluate evidence properly.[2]

What about reasons that are neither epistemic nor moral? Say, in particular, that we perform a financial cost-benefit analysis, and it turns out that the costs outweigh the benefits. Does that provide a reason against interference? It is not clear that it does in the kinds of situations that concern us in the present case for epistemic paternalism. The reason is that we have moral reasons for interference, as we saw above, and that moral reasons *silence* countervailing, non-moral reasons.[3] For present purposes, such silencing doesn't require *authoritative* moral reasons, operating independently of desires or interests. After all, in so far as we are inclined to interfere with the inquiry of another in the kind of contexts that have concerned us, our motivations are most plausibly understood as grounded in a moral concern, such as a moral concern for the welfare of the defendant or for doing right by those wronged, which silences reasons owing to non-moral considerations. That much should be largely uncontroversial. Those critical of "Humean" silencing (e.g., Joyce 2006) are not concerned that moral reasons grounded in *present* desires cannot silence non-moral reasons—they are concerned that this is the only

kind of silencing there is, on a Humean picture. But that complaint need not concern us here. That means that, if we can show that the epistemic reasons involved are aligned with our moral reasons on the issue, then no further kinds of reasons need to be considered for us to conclude that the alignment condition is satisfied. Consequently, the aforementioned type of evidence control can be taken to satisfy that condition.

6. Conclusion

We started out by arguing that evidence regarding our dual tendency for bias and overconfidence suggests that our best bet when it comes to counteracting bias and promoting epistemic goods is to have external constraints imposed that restrict our freedom to conduct inquiry in whatever way we see fit. Moreover, it was argued that practices that impose such constraints are properly referred to as epistemically paternalistic when they interfere with our freedom to conduct inquiry in whatever way we see fit, and do so for our own epistemic good without consulting us on the issue. Two objections to such interference framed in terms autonomy violations were rebutted, and two jointly sufficient conditions for justified epistemic paternalism then defended. Finally, it was argued that the practice of evidence control used throughout the chapter to illustrate the idea of epistemic paternalism satisfies those conditions. As it happens, there are other, strong candidates for justified epistemic paternalism, including the practices of relying on experimental randomization in the sciences, and on statistical prediction rules in medical diagnosis and prognosis.[4] But for present purposes, the case of evidence control suffices to demonstrate that we are sometimes justified in interfering with the inquiry of another without her consent but for her own good, and that epistemic paternalism consequently is true.[5]

References

Ahlstrom-Vij, K. (2013*a*) 'In Defense of Veritistic Value Monism,' *Pacific Philosophical Quarterly*, 94(1), 19-40.

Ahlstrom-Vij, K. (2013*b*) *Epistemic Paternalism: A Defence*. Basingstoke: Palgrave Macmillan.

Alicke, M. D. (1985) 'Global Self-Evaluation as Determined by the Desirability and Controllability of Trait Adjectives', *Journal of Personality and Social Psychology*, 49, 1621-30.

Borgida, E., and Nisbett, R. (1977) 'The Differential Impact of Abstract Vs. Concrete Information On Decisions', *Journal of Applied Social Psychology*, 7 (3), 258-71.

Brown, J. D. (1986) 'Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments', *Social Cognition*, 4, 353–75.

Cosmides, L., and Tooby, J., (1996) 'Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty', *Cognition* 58, 1-73.

de Marneffe, P. (2006) 'Avoiding Paternalism', *Philosophy & Public Affairs*, 34, 68-94.

de Marneffe, P. (2010) *Liberalism and Prostitution*. Oxford: Oxford University Press.

Doris, J. M. (2002) *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Dworkin, G. (2010) 'Paternalism' in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*; available at http://plato.stanford.edu/archives/sum2010/entries/paternalism/.

Feinberg, J. (1986) *The Moral Limits of Criminal Law, Volume Three: Harm to Self*. New York and Oxford: Oxford University Press.

Foot, P. (1978) 'Are Moral Considerations Overriding?' in her *Virtues and Vices*. Oxford: Oxford University Press: 181-8.

Gigerenzer, G., Todd, P. M., and the ABC Research Group (eds) (1999) *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press.

Gilovich, T., Griffin, D., and Kahneman, D. (eds) (2000) *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.

Goldman, A. (1991) 'Epistemic Paternalism: Communication Control in Law and Society', *The Journal of Philosophy*, 88 (3), 113-31.

Goldman, A. (1992) 'Foundations of Social Epistemics' in his *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (pp. 179-207). Cambridge, MA and London, England: The MIT Press.

Hart, H. L. A. (1963) *Law, Liberty, and Morality*, Stanford, CA: Stanford University Press.

Joyce, R. (2006) *The Evolution of Morality*. Cambridge, MA and London: The MIT Press.

Kahneman, D. (2011) *Thinking, Fast and Slow*. London: Penguin Books.

Kahneman, D., Slovic, P., and Tversky, A. (eds) (1982) *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, MA: Cambridge University Press.

Laudan, L. (2006). *Truth, Error, and Criminal Law*. Cambridge: Cambridge University Press.

Lupfer, M. B., Weeks, M., and Dupuis, S. (2000) 'How Pervasive is the Negativity Bias in Judgments Based on Character Appraisal?' *Personality and Social Psychology Bulletin*, 26, 1353-66.

Maeder, E., and Hunt, J. (2011) 'Talking about a Black Man: The Influence of Defendant and Character Witness Race on Jurors' Use of Character Evidence', *Behavioral Sciences and the Law*, 29, 608-20.

Mill, J. S. (1989) 'On Liberty', in S. Collini (ed.) *On Liberty and Other Writings* (pp. 1-116). Cambridge: Cambridge University Press; originally published in 1859.

Petty, R., and Wegener, D. (1993) 'Flexible Correction Processes in Social Judgment: Correcting for Context-Induced Contrast', *Journal of Experimental Social Psychology*, 29, 137-65.

Pronin, E. (2007) 'Perception and Misperception of Bias in Human Judgment', *Trends in Cognitive Science*, 11 (1), 37-43.

Pronin, E., Lin, D., and Ross, L. (2002) 'The Bias Blind Spot: Perceptions of Bias in Self Versus Others', *Personality and Social Psychology Bulletin*, 28, 369-81.

Raz, J. (1986) *The Morality of Freedom*. Oxford: Clarendon Press.

Shiffrin, S. V. (2000) 'Paternalism, Unconscionability Doctrine, and Accommodation', *Philosophy and Public Affairs*, 29 (3), 205-50.

Taylor, S. E., and Brown, J. D. (1988) 'Illusion and Well-being: A Social Psychological Perspective on Mental Health', *Psychological Bulletin*, 103, 193-210.

Wilson, T. D. (2002) *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.

Wilson, T. D., and Brekke, N. (1994) 'Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations', *Psychological Bulletin*, 116 (1), 117-42.

Wilson, T. D., Centerbar, D. B., and Brekke, N. (2002) 'Mental Contamination and the Debiasing Problem', in Gilovich, Griffin, and Kahneman (2002), pp. 185-200.

Young, R. (2008) 'John Stuart Mill, Ronald Dworkin, and Paternalism', in C. L. Ten (ed.) *Mill's 'On Liberty': A Critical Guide* (pp. 209-27). Cambridge: Cambridge University Press.

Further reading

No one interested in paternalism of any kind can ignore J. Feinberg, *The Moral Limits of Criminal Law, Volume Three: Harm to Self* (New York and Oxford: Oxford University Press, 1986). A more recent and also very insightful discussion of paternalism is P. de Marneffe, *Liberalism*

and Prostitution (Oxford: Oxford University Press, 2010). K. Ahlstrom-Vij, *Epistemic Paternalism: A Defence* (Basingstoke: Palgrave Macmillan, 2013) makes a case for the type of epistemic paternalism outlined in this chapter, and defends a variety of epistemically paternalistic practices. A. Goldman, 'Epistemic Paternalism: Communication Control in Law and Society', *The Journal of Philosophy*, 88 (3) (1991): 113-31, introduced the term 'epistemic paternalism', and contains a discussion of some epistemically paternalistic practices, including that of evidence control.

Notes

[1] That the judge is interfering within the *legitimate* domain of the jurors might be taken to suggest that her action is at least *pro tanto* morally objectionable (see, e.g., Shiffrin 2000: p. 220, footnote 25). However, whether it's all-things-considered morally objectionable would depend on what other factors are in play. Suffice for present purposes to note that, in the type of case at hand, any legitimate moral claim on the part of the jury not to be interfered with is trumped by considerations about the welfare of others, such as the defendant and those allegedly wronged by the defendant. For more on the moral reasons at work here, see Section 5.

[2] What about the fact that the judge is arguably interfering within the legitimate domain of the jurors? See note 1 above.

[3] What about situations involving *immense* financial costs? In such situations, we might hesitate to say that moral reasons silence whatever countervailing reasons we might have (see, e.g., Foot 1978). For present purposes, however, it suffices to note that the kind of situations that will concern us do not involve any such immense costs.

[4] See Ahlstrom-Vij (2013*b*) for an argument to this effect.

[5] Parts of the present chapter are reproduced from Ahlstrom-Vij (2013*b*) with the permission of Palgrave Macmillan.