Forthcoming in Ian J. Kidd, Heather Battaly, Quassim Cassam eds. *Epistemic Vice* (Routledge)

Institutional Epistemic Vices: The Case of Inferential Inertia

One of the achievements of virtue epistemology has been the identification of an array of epistemic virtues that had not previously been distinguished or focused on in traditional epistemology.[1] Similarly, it is an ongoing achievement of what Quassim Cassam has called 'vice epistemology' to explore the underbelly of the same domain—the epistemic vices that either mirror or in less direct ways reflect the constellation of virtues. There is also room for a certain hybridity across the two main domains in which the notions of virtue and vice find application—the epistemic and the ethical. In earlier work I tried to bring attention to two virtues that were hybrid in that they are both epistemic and ethical in kind, since they aimed equally at the ultimate ends of truth and justice. These were 'testimonial justice' and 'hermeneutical justice' (Fricker 2007), both virtues of epistemic justice; and I subsequently tried to show how these virtues might manifest themselves not only as individual virtues but alternatively as institutional virtues, gatekeeping certain kinds of political power (Fricker 2010 & 2013). José Medina has explored further virtues and vices in the domain of epistemic injustice—virtues such as humility, curiosity/diligence, and open-mindedness; vices such as epistemic laziness, arrogance, and closed-mindedness (Medina 2013)—and, I take it, all such virtues and vices will be hybrids inasmuch as the wrongs that they pre-empt or that flow from them respectively are at once ethical and epistemic in kind. Just as there can be hybrid ethical-epistemic virtues, then, so there can be hybrid ethical-epistemic vices.

Why might we care about the question whether institutions can be said to have vices of any kind? Why not content ourselves with using a vocabulary of, say, efficiency and inefficiency,

functionality and dysfunctionality, relative to the institution's goals or purposes? Surely we are able to critically assess these things, and the relative merits of the institutional goals and purposes too, without having to talk specifically in terms of virtue and vice?—terms which after all strike many as having alienating overtones of high church or, alternatively, high classicism, depending on whether they ring moralistic or simply archaic. This is a fair question; but on the other hand, let's not forget that (to fleetingly sloganize) meaning is use—or at any rate these concepts and their overtones are not static, unless and until we stop using them. There might be good reasons to rehabilitate and normalize the notions of virtue and vice, even for institutions, so that unwanted overtones are silenced and ideas of virtues and vices come, more completely, to seem like a proper part of our contemporary normative equipment for ethical evaluation. Indeed this is one way of picturing what much of our recent theorizing about virtue and vice is gradually working to achieve.[1]

More specifically, an important reason to engage in the ongoing modernization of these concepts is that something approximating the idea of institutional vice has for some time had a tentative foothold in public discourse, but one we perhaps do not yet conceive very clearly either in the public domain or in philosophy. The idea of institutional vice, or virtue for that matter, has even now received relatively little philosophical attention compared with individual virtue and vice, so perhaps there is some useful work to be done on this score.[2] A central role for the idea of institutional vice came sharply to the fore in British culture in 1999 with the publication of the Macpherson Report on the handling of the racially motivated murder six years earlier of a teenager, Stephen Lawrence, in South East London (Macpherson, 1999). That report described

[1] For a defence of the need, indeed 'inevitability', of such virtues see Madva (2019).

the London Metropolitan Police as 'institutionally racist', and this marked a watershed moment of public acknowledgement of the deep permeation of racism in a central and powerful institution such as the capital's police force.[3] If anything is a vice in an institution then racism is, and I therefore consider the idea of a police force being found to be institutionally racist as furnishing us with a central and prominent example of something that can be properly theorized as *institutional vice*. While I would not pretend that the idiom of institutional *vice* is the only one in which we could make good sense of the various phenomena of institutional racism (we could, for instance, restrict ourselves to talk of institutional prejudice, inequality, bias, dysfunctionality, failure of protocol…) still I would argue that there is a robust purpose for which the vocabulary of institutional virtue and vice is distinctively well placed to serve. That purpose is basically one of picking out aspects of institutions that are the collective analogue of an individual agent's character, but where the actual individuals whose combined epistemic agency comprises the institution's epistemic agency need not, as individuals, have any of the traits or attitudes of the institution. While some institutional vices will depend upon some significant number of the individual officers having the vice themselves, at least when in role as an officer of that institution, other vices will be more structural in kind, and the notion of institutional vice is well-designed to be applied to both sorts of case. Or so I aim to show.

Ethos Matters

What is distinctive about the idea of an analogue to an individual's character? Why put oneself to all the philosophical trouble of substantiating the idea of *institutional character*? Even if it is philosophically do-able, do we really need it? The answer lies in how far we value the possibility that (at least some of) our institutions have an *ethos* from which their procedures and judgements

flow.[4] Only through sustaining an ethos that guides and explains their conduct might an institution—the NHS, parliament, the police force, the BBC, the care system, the judiciary—genuinely *stand for something*, and constitute part of the fabric of what we believe or hope is good about the culture of which they are a part. Most saliently perhaps, in a democracy it may be important to us as citizens that the judicial system operate not only in a way that is well-designed to deliver right results—fair sentencing in the criminal courts, for instance—but moreover that the institutional mechanisms and procedures that furnish these right results are fuelled by appropriate values. Were the right results (the fair sentencing) produced by a miracle of clever incentivization and efficiency mechanisms, this would not be enough. While there is a place for incentivization and efficiency mechanisms in any institution—performance reviews, prospects of promotion, disbarring and dismissal for anyone found to be corrupt, and so on—the point remains that no institution can produce justice proper except by way of *reasons of justice* (Fricker 2013). In the institutional setting this can only be a matter of the value commitments being embodied in the processes, including the epistemic processes involved in fact-finding and adjudicating, by way of an appropriate *ethos*. Ethos, at least as I am using the notion, looks to be the only way that institutional bodies can incorporate intrinsic values in their agency as institutions. It is comprised of collective motivational dispositions and evaluative attitudes within the institutional body, of which the various good or bad ends orientate the institution's activities.

A different aspect of the explanation why we regard an appropriate ethos, and not merely appropriate outcomes, as important for at least some of our institutions is that a chief way of convincing people that a given outcome *is* appropriate—fair, or just, or a correct application of the rules—is to show them the appropriate value commitments that were in fact brought to bear

in the process that produced it. A good deal of our confidence in institutional judgements and actions flows from our degree of confidence in its value commitments—epistemic values such as actually caring about the truth and the gathering of proper evidence, as opposed to just securing a conviction, for instance. That is why transparency in many of these processes is a good thing—we get to see the ethos of the institution laid more or less bare in the record (video of interviews, Hansard, Minutes of meetings etc.). If, for example, a local government body entrusted with certain town planning decisions in an area of redevelopment fails to consult long-term local residents, or consults only private residents and ignores council housing residents, then we might well describe that institutional body as 'high-handed', 'arrogant', perhaps 'cynical', not only ethically but intellectually—from the point of view of data gathering and achieving a proper perspective for judgement. In itself the outcome judgement (a new supermarket and fitness centre in place of the beloved but little-used Victorian municipal swimming pool) might be found by all to be the right result all things considered, yet still if residents are signally unimpressed with the values implicit in the *way* the authorities came to the planning decision, then this will tend to cast doubt on the quality of the outcome judgement itself. Institutional ethos matters, then, partly because the presence or absence of appropriate value priorities behind any given item or process of institutional epistemic conduct is a factor in determining confidence and satisfaction levels in the outcome judgement itself. Just as value-dispositions matter in our evaluations of individual agents, then, in many of our evaluations of institutional agency, ethos matters.

Insofar as this is a key rationale for modernizing the ideas of virtue and vice with a view to rendering them fully applicable to contemporary institutions, it is worth noticing that it entails a

commitment to a broadly (though not exclusively) motivational or disposition-based conception of virtue/vice. Were one to adopt instead a purely skill-based conception that makes no mention of the agent's motivational states and dispositions, then there would be no distinctive purpose for virtue and vice talk in relation to institutions. We might just as well stick with the familiar, thinly performance-oriented terms of assessment, evaluating institutional epistemic agency in relation to informational accuracy, size of data sets, evidential thoroughness, breadth of fact-finding, soundness of predictions, and other aspects of epistemic performance construed in a way that does not draw on ethos. So the general question 'What is distinctive about the idiom of virtue and vice when it comes to evaluating institutional conduct?' has furnished an answer that makes a purely skill-based conception otiose. I shall therefore pursue my line of thought about institutional vice on the basis that epistemic virtues involve good epistemic dispositions and attitudes as well as reliability in achieving good epistemic ends. This means I will be using a broadly responsibilist conception as opposed to a reliabilist conception (Axtell 2000; Battaly 2010).[5]

In saying we care, and should care, about an institution's ethos, I have so far put the point only positively: we want justice proper, and so we want the judiciary to produce just outcomes from an appropriately stable commitment to the value of justice—an ethos of justice. However, as soon as one puts the point positively in this way, the negative counterpart quickly comes to mind: the idea of institutional ethos is equally important because we need to be able to think about our institutions critically, in terms of their faults, whether stable and systematic, or fleeting and one-off. Like individual agents, institutional bodies can obviously have fleeting lapses of judgement that might even be described as 'out of character'. That is why it is always such an

important question to ask of an apparently one-off lapse whether it is indeed fleeting and out of character or whether in fact, beneath the surface, there has been a deterioration of ethos more widely in a branch of the organization. This question was in the air, for example, when on the 9th February 2018 it was revealed by *The Times* that 'Top Oxfam staff paid Haiti survivors for sex', after which a crucial question was whether similar abuses had been committed by other Oxfam aid workers, and whether Oxfam had in any way covered up allegations or important details of the case. Assessing how far a given lapse is a one-off event or an expression of more systemic decline in institutional ethos is an important question, partly because it determines whether or not a general loss of faith in the organization is a warranted response, and because of the implications for what it will take to fix the organization.

We need, then, to be able to make reference to institutional ethos when engaging in evaluative judgements about institutions' practical and epistemic functioning. And this means we have an interest in understanding what social-metaphysical commitments are entailed by such talk. What, then, is the structure of an institutional ethos?

Modelling Ethos

I am proposing the idea of an ethos as the institutional analogue of an individual's character in the virtue-theoretic sense—a set of interrelated dispositions and attitudes, where (in the case of a virtuous person) these are conceived as temporally and counter-factually stable motives towards good ultimate and mediate ends. This will be the same whether we are considering ethical or intellectual character, though the good end(s) are fewer in the case of the intellectual. (Value

monists would say truth is the only ultimate end of epistemic virtue, so that different epistemic virtues are exclusively individuated by reference to their differing mediate ends; pluralists, by contrast, allow that an intellectual virtue might aim variously at truth or knowledge or understanding... Zagzebski helpfully glosses these by talking in terms of 'cognitive contact with reality'.) In modelling the phenomenon of institutional ethos, there are two broad theoretical approaches available to us that are applied to various sorts of group phenomena, such as group intentionality, group belief, and so on: summative and collective. In general, it should not be thought of as a competition, for both models represent perfectly real possibilities, and indeed each kind of group is frequently realized in institutional life. For example, if an exam board were to move by way of majority vote to produce the judgement that a given candidate has satisfied the requirements for conferral of degree, the group and its decision is structured summatively. If a similar exam board in a neighbouring department moves by way of consensus towards what Margaret Gilbert calls a 'joint commitment' to the judgement that a given candidate has satisfied the requirements for conferral of degree, then the group is structured as a collective, and in the committed sense she calls a 'plural subject'. In this latter case the group judges 'as a body', to use Gilbert's evocative phrase (Gilbert, 2000). This will involve the individual board members expressing willingness, under conditions of common knowledge (each knows that the others know….), to endorse or go along with the judgement.[6]

By contrast, the summative model of institutional ethos would cast institutions, or institutional bodies within broader institutions (faculties within a university, or squads within a police force) as capable of having an ethos only as a matter of the aggregate of individual officers' value commitments. These may or may not be adequately consistent to add up to a coherent ethos, and

so for any such aggregation of individual values it will remain an open question whether they amass into a coherent ethos or not. But it is worth noting that a further possibility for institutional ethos is a mixed economy of collective and summative structures. If, for instance, members of the executive branch of an institution were to jointly commit to a certain set of values in their deliberations and judgements, while the implementation branches of the same organization were charged with simply implementing the policies (without any joint commitment, let us imagine, to the values from which they flowed), then there could easily be some significant mismatch between the jointly committed values of the executive and the values of the officers implementing them on the ground. They may or may not like it but they do it regardless as part of their job because that's the new policy they have been told to implement. Notably, in a case where the executive body is trying to bring about institutional *change*, we must positively expect such mismatch. This mixed economy picture presents us with a fairly typical top-down form of institutional change (or failure to change, depending on how it turns out); and one of the reasons why top-down change can ultimately fail is precisely because successful institutional change so often depends on the new ethos being genuinely taken up at the implementation level of the organization—the officers in the field, whether policing, or healthcare, or development work. While there is room for a mismatch between newly committed values in the executive and their mere implementation at the ground-level, then, in many situations (recall the case of Oxfam) if those values are not in fact stably held by the officers in the field, whether by dint of joint commitment or personal commitment, then there may be trouble ahead.

Anyone looking to make good sense of institutional ethos will need to make room for non-summative possibilities, including the mixed economy model above. I have started with Gilbert's

view, and will argue that it offers the appropriate collectivist model for present purposes; but we should first acknowledge that there is a range of differing possibilities of a genuinely collectivist kind. The various options can all be helpfully conceived in relation to the degree of *commitment* to the shared intention/value/activity, as it may be. Michael Bratman's view of collective intentionality requires no *commitment* as such, in the sense that the interdependent intentions of members of the group who each intend to play their part in a planned group activity, such as painting a house together, are strictly *pro tem*, so that someone who breaks away is not burdened by any residual, commitment they are failing to honour (Bratman, 1999 & 2014). Christian List and Philip Pettit (2011) do not employ any stronger idea of commitment to group plan than we find in Bratman; indeed they are explicit in using more or less his conception. Raimo Tuomela's conception of 'we-thinking' does involve a certain commitment to the group endeavour, but he employs the notion of 'ethos' to capture the flavour of that commitment, and it does not seem to entail that a break-away member of a we-intending group would thereby be the proper object of rebuke (Tuomela, 2013). If this broad contrast between Gilbert's view and the others in the literature is correct, then Gilbert's account emerges as distinctive in embedding a strong notion of commitment in the very mechanism of group agency—the formation of the plural subject. The joint commitment involved in the making of any collective judgement—for instance, the examining board's judgement that the examinee has satisfied the requirements for conferral of degree—survives even the most complete lapse of intention, participation, or interest on the part of a given individual member of the group. For Gilbert, if a member reneges on a joint commitment, they are therein a proper object of rebuke or a demand for an explanation.

One can become fully party to a joint commitment of this kind even if one personally disagrees with the content of the commitment, for one can go along with a given judgement that the candidate has satisfied the requirements for conferral of degree by merely 'letting it stand', even if, as a private individual you would not make that judgement. Perhaps you take a dim view of one of their exam results, considering conflicting examiners' marks to have been resolved in the wrong direction, but because your colleagues on the board take a different view, you have decided to acquiesce and allow a consensus. Going along with a judgement by 'letting it stand' in this way is sufficient for being fully party to the jointly committed judgement. And it is perhaps worth noting that this feature of the view is highly desirable, because it is what enables boards and committees and other institutional bodies to achieve unity in group judgements even while there may remain candid disagreement at the individual level. This is how Gilbert's model works to allow potentially radical differences between a judgement of the collective body and the judgements of individuals that compose it (Gilbert 1987 & 2002). Still, once an individual group member is party to such a commitment, they are on the hook.

This binding aspect of Gilbert's plural subjectivity naturally renders it somewhat susceptible to objection when the model is applied to the breezier forms of group activity that seem more easy-come-easy-go. If people dance together for a while in a salsa club, and then one of them has had enough and wanders off to get a drink, isn't that perfectly fine? No need for explanation, surely, let alone rebuke—that would be weird. *That* kind of dancing together is non-committal, but it is still *dancing together*. It lasts as long as it lasts, and that is part of the freedom of it. Bratman and Tuomela's models can easily accommodate this kind of case, though obviously not Gilbert's which is too demanding on the commitment front. But here I believe the correct conclusion to

draw is that a certain pluralism is in order to accommodate the full range of ways we can do things together. It is precisely the bindingness of Gilbert's model that makes it the right one for present purposes, for in order to make sense of the bindingness of ethos we *need* the commitment involved in the joint commitment that creates the plural subject of the ethos. An ethos is precisely not something *pro tem*, but something committed and intersubjectively binding by way of potential apt rebuke. A set of values that one can ditch when it no longer suits is no ethos at all, but mere lip service. The ethos of a group or institutional body is something that binds its members because it consists of value commitments in the robust sense of commitments that are temporally and counterfactually stable, or at least *meant to be* temporally and counter-factually stable. If a member proves their commitment less than appropriately stable by reneging on it, then they *are* properly subject to rebuke. I conclude that the joint commitment model is the distinctively appropriate model to employ in elaborating the collective value dispositions involved in forming a given institutional ethos, and therefore the appropriate model in particular for the idea of an institutional epistemic ethos.

To use the Gilbertian apparatus for these purposes is basically to engage in a piece of analogical thinking. Indeed all talk of plural subjectivity is based on the elaboration of an analogy between the individual level and the group level, so that we may earn the right to speak literally of groups doing things like making a careful judgement that a given candidate has satisfied all the requirements for conferral of degree. Gilbert explains the very idea of a plural subject by way of an analogy between the relation of an individual's will and the action it produces with the relation of a group's pooled wills and the relation with the group action produced: 'these wills will be directed at that end, *as if* they belonged to a single person. That is, the coherence of the

behaviour which is their output will approximate in coherence to the output of the will of a single person acting in pursuit of a goal of his own' (Gilbert, 1989; p. 211). For our distinctly epistemic purposes at the level of institutions, my suggestion is that we invoke a similar analogy. We say that when an institutional body, like an examinations board, comes to a careful judgement that the candidate has satisfied the requirements, the board members' individual wills (in the aspect relevant to epistemic agency) are directed at the end of coming to a careful judgement *as if* they belonged to a single adjudicator. On this basis, we may speak simply and literally of *the exam board* making the careful judgement in question.

As we have seen, this model requires individual members to express willingness, under conditions of common knowledge, to at least go along with the group judgement made. Transposing this now to our proposed case of institutional epistemic ethos—an ethos, for example, of truthfulness and fact-checking in public office—we might say of an institutional body such as a branch of government: the individual officers all endorse, or at least go along with, the set of values that comprise an ethos of truthfulness and fact-checking in public office. Thus parties to this jointly committed ethos would in theory rebuke anyone who was found to have lied or bullshitted in public office. It is worth emphasizing, as Gilbert does, that the expression of willingness can be very minimal. Indeed in some contexts one can surely count as expressing willingness by default, if you simply fail to object. Imagine, for example, a case in which there is general suspicion of a figure in political office that he is using his influence for personal political gain in an upcoming election—and yet no one is yet talking about it openly, and there remains a kind of group pretence that everything is as it should be, or at least within the bandwidth of normal political dealings. Those officers of the political institution in question are

in the unfortunate position that their silence is a way of going along with the fiction that everything is okay. But it isn't; and in their continued silence they passively become party to a joint commitment whose content is incompatible with the good institutional epistemic ethos proper to a democratic government. It will take a whistle-blower to break the silence, and that is a seriously costly thing to be for all sorts of reasons, but one feature of the pressures on a potential whistle-blower is that they are currently party to a joint commitment of silence. Whistle-blowers are admired by many, but they are also rebuked and often abused by many who wish to discredit them. Much of that is more politics and sheer threat in the context of power-mongering; but one aspect of it is proper to the basic normativity of joint commitment. It is built into Gilbert's model that if you do at least go along with the silence for a while, you thereby become party to the joint commitment to conserve the status quo, so that if you then shift your stance and blow the whistle you are *pro tanto* a proper object of rebuke. In a real political case that may well be the least of an actual whistle-blower's worries, but it is a real normative feature of such situations none the less, and one that can be exploited by those with an interest in conserving the silence.

As Gilbert herself has emphasized in relation to group belief, this brings out the insidious potential of joint commitment—once made, it brings genuine normative pressure to bear, and there is no guarantee that this is to a good end of any kind. As in the example just given, it might work in the interests of a corrupt politician by creating genuine normative pressure not to break ranks. We can see better just how easily—how *passively*—this kind of situation can be manufactured if we reflect, further, on the phenomenon of 'accommodation' as Rae Langton has recently explored it. She takes up David Lewis's notion of accommodation as a feature of

scorekeeping in a language game (Lewis 1979). Lewis coins the idea to capture how a presupposition can enter circulation in a conversation and be accommodated if it is not actively challenged; and Langton elaborates how powerful this mechanism can be as a means of introducing prejudiced, stereotypical or hateful ideas into conversational circulation. She calls assertions that enter in by way of presupposition 'backdoor testimony'—assertions, indeed tellings of an inexpicit kind, that surreptitiously become accommodated and thereby perpetuated without challenge (Langton 2018). Similarly, I would suggest that the phenomenon of accommodation is useful for seeing one way that a default 'expression of willingness' to go along with the content of a given joint commitment can be a powerful force. An official can passively become party to a joint commitment to keep quiet about a political leader's corrupt lies just by failing to dare to be a whistle-blower. One thereby accommodates the presupposition that the leader had engaged merely in acceptable levels of political hyperbole or bombast, even if one knows perfectly well they were lying. In such a case, a person with decent epistemic values becomes party to a committed toleration of corrupt mendacity, thereby passively betraying whatever may be left of the decent epistemic ethos of truthfulness in public office.

Becoming party to a joint commitment, then, is easy—frighteningly so in some contexts—and this can be manipulated by those who wish to keep certain attitudes in circulation. The net result at group level is that the institutional body in question behaves in a way that departs from a pre-existing ethos of truthfulness and fact-checking. In this respect Gilbert's model emerges once again as an excellent fit for modelling institutional ethos; only now we are concentrating on its credentials for modelling bad or vicious ethos. Every time officers in the government of our imagined truthless political leader go along with something he has said or presupposed, passively

letting it stand and thereby accommodating it, they raise the level of pressure—genuine normative pressure—to conserve the status quo. The mechanism of joint commitment helps us understand a normative aspect of conspiracies of silence; but more importantly for present purposes it lays bare the mechanism of how institutional bodies can behave in ways that depart from, and help deteriorate, a pre-existing epistemic ethos of truthfulness and fact-checking. The insidious joint commitment to letting stand and thereby increasingly accommodating truthless content threatens to end in institutional epistemic failure—shifts of enduring commitment that risk entailing substantial erosion of good epistemic ethos in the medium to long term, bringing about a corruption of institutional character.

Modelling Institutional Epistemic Vice

Epistemic vices, like vices in general, can pertain to acts or behaviours, and they can pertain to agents. When a given agent displays a stable pattern of vicious actions and behaviours, then we attribute the vice not only to that which is done but also to the doer themselves. I committed myself earlier to a broadly responsibilist conception of virtuous action whose distinctive feature is its motivational richness—an inner or motivational element. No other model would be able to capture the importance we implicitly attach to the ethos at the heart of certain of our institutions. But it also includes a reliabilist aspect—an outer, or performative element. On the responsibilist conception, an epistemically virtuous agent is someone who acts from a temporally and counter-factually stable good motive *and* where the good end of that motive is reliably achieved (Battaly 2010; Zagzebski 1996). Thus we are presented with both an inner and an outer element of virtue, and therefore two distinct areas for potential failure and lapse into vice.

But not all failures will indicate vice. There must surely be some more neutral ground, not least because virtue itself comes in degrees, as an agent is increasingly habituated and spontaneous in her responses. If an agent—individual or institutional—acts in a manner that is less than ideal, but not culpably bad, then it would not be natural to use the word 'vice' to describe either them or their action. Imagine, for instance, an institution that has a good ethos, and so the inner element is fine, yet is inefficient in its performative aspect, so the outer aspect is less than it might be. A school, perhaps, with teachers who care about doing a good job, but an IT system for homework submission and marking that is not well managed or well used, and makes for occasionally serious communicative failures both with students and with parents. This is not an institution we would hold up as an exemplar of epistemic virtue as regards its information sharing practices; indeed we would criticize it, but we would not go so far as to describe it as guilty of any epistemic vice.[7]

This said, we can return to the previous point that the responsibilist conception makes for two distinct areas of potential culpable failure—the inner ethos (stable motives), and the outer performance (achievement of the ends of those motives). If vices are culpable failures of virtue, then epistemic vices are culpable failures of epistemic virtue either in respect of ethos and/or in respect of (what we might call) implementation. Imagine our school does seriously mess things up one year, so that teachers have practically given up using the online homework system but no proper alternative has been put in place. And imagine the mess up is bad enough—disruptive enough to cause a real loss of confidence on the part of the students and parents—so that the school *is* culpable. Now we are looking at a behaviour on the part of the school that would count

as vicious as regards its practices of information sharing. But this 'thinking vice'[8] might yet be out of character for the school, so it does not yet imply that the school itself—the institution—has the vice. That attribution would require a temporally and counter-factually extended pattern of such culpable lapses of virtue as regards information sharing practices. So let us now imagine our school ten years on, after a decade of becoming increasingly inefficient and disorganized, even while it had opportunities to do better. The teachers have become disenchanted and fed up so that despondency and laziness has infected the whole-school ethos; and/or, despite continued underlying value commitments, the school has simply fallen into repeated performative failures in the implementation of its policies on information sharing. One way or the other this would now be a school that displayed an epistemic vice of bad information sharing. The various culpable failures of virtue have congealed into a systematic failure, so that the very character of the institution has been changed for the worse. Institutional epistemic vice is a matter of *culpable epistemic bad habits*, where the culpable lapses might be in ethos or in implementation, or in both.

Charlie Crerar has critiqued what he calls the 'mirror view' of epistemic vices, one form of which depicts vices as always positively aiming at an epistemically bad end, mirroring the way virtues always aim at a good one (Crerar, 2018). The critique is persuasive—in fact it is not easy to dream up even a single psychologically coherent epistemic vice of that kind because of the fundamental investment in truth/knowledge that all epistemic subjects as such have. Jason Baehr imagines a far-fetched case of epistemic malevolence that would fit the bill (Baehr 2010); and perhaps another promising prospect in this regard might be persistent kinds of self-deception—imagine someone with a long-term investment in lying to themselves about how talented they

are, for instance. Still, such cases will surely be unusual at best, so that on the whole any motivational disorder constituting an epistemic vice will instead take the negative form of an inadequate commitment to good epistemic ends. These ends might be the ultimate end of cognitive contact with reality—as in the case where the politician shows a flagrant disregard for the truth—or, alternatively, any of the mediate ends whose epistemic value consists in their functioning as a means to achieving that ultimate end. Such mediate ends might be, for instance, looking carefully at the evidence, fact checking, being open to counter-arguments, realizing when one's evidence base is too narrow, and so on. It may not be possible to be epistemically virtuous without an appropriate commitment to the ultimate end of cognitive contact with reality, but it does not follow that only a lapse in relation to that ultimate end can indicate epistemic vice. On the contrary, I contend that someone might be epistemically vicious precisely because of persistent lapses in relation to a mediate end—such as fact-checking—even if their ultimate epistemic commitment remained intact. This possibility is entirely compatible with the background theoretical idea that only the value of the ultimate end confers value on the mediate end—so, for instance, the only reason fact-checking matters epistemically is because fact-checking promotes cognitive contact with reality. Indeed my contention positively relies on the instrumental connection between mediate and ultimate ends, for what makes it epistemically bad, and potentially culpable, to fail to fact-check is precisely that failing to fact-check is bad from the point of view of achieving cognitive contact with reality. It is because of this instrumental connection between fact-checking and cognitive contact with reality that a lapse in the former can, if culpable, constitute an epistemic vice. As regards this kind of epistemic vice, the distinguishing feature is that the subject is to blame for how at least one of their mediate motives is failing to align with the ultimate end of cognitive contact with reality.

Crerar imagines two figures, Galileo and Dave, each of whom seems to present a clear case of epistemic vice and yet each of whom is equally clearly committed to epistemically good ultimate ends.[9] Galileo is individually brilliant and cares about the truth, but he is also epistemically arrogant in his scornful neglect of the views of his colleagues; Dave displays a lamentable narrow-mindedness of the privileged, even though he too cares about the truth. I agree these characters surely display epistemic vices, and I agree that they would represent a puzzle to anyone arguing for a conception of epistemic vice that required motivational lapse in respect of ultimate ends. But for the reasons offered above, I would interpret Galileo and Dave as exemplifying epistemic vices in virtue of the fact that each is culpably unmotivated towards a relevant *mediate* end—respectively, that of listening to one's fellow researchers' informed opinions, and that of awareness of how privilege is affecting one's social perceptions. Such culpable lapses in relation to mediate epistemic ends is perfectly sufficient for epistemic vice.[10] Like Crerar I do not hold to any exclusively motivational conception of vice, but this is not because I do not see Galileo and Dave's epistemic failings as motivational failings. I interpret both Galileo and Dave as displaying vices by exhibiting mediate motivational failings that undermine their epistemic orientation to cognitive contact with reality, even while they both may remain psychologically motivated to achieve it. For me the reason to reject an exclusively motivational account of vice is simply that motivational failure is not the only route to vice, since a culpable lapse in the outer, performative aspect of virtue remains an independent possibility. Remember our informationally challenged school at the moment where its flawed information-sharing practices have become seriously entrenched, despite a continuing good ethos. This school is displaying an epistemic vice of sloppy information sharing even though there is nothing

wrong with it at the motivational level of ethos; the problem and the culpability is all at the performative level. A person or institution can display epistemic vice simply through persistent performative failure, even if the motivational commitments, mediate and ultimate, are all that they should be.

On the view I am putting forward, then, epistemic vice consists in some culpable lapse of epistemic virtue either (i) in its inner aspect of mediate and/or ultimate motivations to good epistemic ends, and/or (ii) in its outer aspect of performance—the achievement of those ends. A motivational and/or performative lapse that is bad enough to warrant blame is bad enough to warrant the label 'vice'. Where it is persistent it will constitute a vice of epistemic character and not merely a more fleeting vice of thinking. Putting together the earlier picture of institutional ethos with this conception of epistemic vice, we can say that *institutional epistemic vices are displayed—either in thinking or, where persistent, also at the level of institutional character—whenever there are culpable lapses in the institution's epistemic ethos and/or in the implementation of its ends*.

The Institutional Vice of Inferential Inertia

A salient rationale for a philosophical vindication of the idea of institutional epistemic vice is that there may be some epistemic vices that are especially worth identifying in their institutional form, either because they are especially pernicious in that form, or because they are especially fixable, or both. I want to draw attention to an epistemic vice I will call the vice of Inferential Inertia. I think we easily recognize a certain scenario in an individual hearer, who is not guilty of perpetrating any testimonial injustice exactly, for their credibility judgement of the speaker is not

depressed by prejudice of any kind, and indeed (let us imagine) they do believe her; and yet…nothing else happens by way of epistemic follow-through. Imagine a case of someone telling a colleague or co-worker of a crime committed in the workplace. Imagine the colleague believes what she tells them – they assent, they express genuine sympathy or shock, or whatever is in order – and yet…somehow the evidential bearing of what they have been told does not impact anywhere (else) in their belief system. Perhaps they are resistant to the implications of this particular piece of news, and hope it is a one-off; or perhaps they are epistemically lazy or scared or unimaginative when it comes to shaking their sense of the *status quo*. Such a person, let us imagine, fails to draw any inferences at all, does not alter her other beliefs one iota, even though they are likely seriously undermined by what she has been told and now believes. For instance, their other beliefs about the perpetrator remain unaltered, or at least are certainly not altered in a manner appropriate to the evidence.[11] They somehow hold the contradictory beliefs in suspension without making the cognitive effort to draw the inferences that are there to be drawn, even tentatively. This person believes what they are told, but the new information never gets to have its evidential impact. They are guilty of a distinctive epistemic vice, that of Inferential Inertia. In such a case, though the speaker is not misjudged epistemically, still she is just as frustrated in her aim to bring the hearer to appreciate the implications of what she's saying as she would have been in an ordinary case of testimonial injustice. From the point of view of inferential uptake, she might just as well have not been believed. We might slot this phenomenon of inferential inertia into relation with testimonial injustice by saying that insofar as any case of inferential inertia is the product of prejudice, then it is a close relation of testimonial injustice, and instantiates a hybrid ethical-epistemic vice; or, alternatively, insofar as a given case is the result of some other kind of epistemically culpable error, it displays a plain epistemic

vice—in this case the kind of stupidity inherent in failing to grasp clear implications of new information.[12]

I think we can easily identify various institutional forms of exactly this epistemic vice. How many feedback forms does one fill out, whether online or by hand, after a doctor's appointment, a retail experience, a meal at a restaurant, an online purchase, or even a trip to the dentist, where the much vaunted feedback, accompanied by apparently sincere declarations of just how much they really want to know how they can do better, in fact passes into an institutional void. Too often the fact of *having* such a mechanism notionally in operation is all that the institution really cares about (they can tick that box), so that the evidential import of any of the content actually fed back is entirely lost. This is the institutional epistemic vice of Inferential Inertia, and it is rapidly becoming a normal part of our institutional environment.

Looking to a gravely serious example from UK institutional life we can see a quite different way in which it can be important to diagnose an institutional epistemic vice: in the BBC commissioned independent review led by Dame Janet Smith into the BBC's culture and practices in the years when Jimmy Savile was committing multiple predatory sexual crimes, we see that this epistemic vice of Inferential Inertia was effectively a key part of the diagnosis of what went wrong institutionally speaking so that he was allowed, even enabled, to commit these crimes undetected for so long.

Dame Janet's review emphasizes certain cultural aspects of the BBC at the time, one of which is a climate of not complaining and in particular not complaining about the Talent:

The Culture of Not Complaining about the Talent

> 54. As I have said, there was a culture of not complaining about anything. The culture of not complaining about a member of the Talent was even stronger. Members of the Talent, such as Savile, were to a real degree, protected from complaint. The first reason for this is because of a deference or even adulation which was, and still can be, accorded to celebrity in our society. The second reason was because of the attitude within the BBC towards the Talent. The evidence I heard suggested that the Talent was treated with kid gloves and rarely challenged. An example of this is the attitude of C51's supervisor when he was told that Savile had sexually assaulted C51 (see paragraphs 5.254-5.255 of my Report). His immediate reply was *'Keep your mouth shut, he is a VIP'* (Smith 2016).

The review also emphasizes a 'culture of separation' and the 'silo mentality' that entailed there was very little information sharing between different parts of the BBC. Competitiveness between departments exacerbated the situation, since it incentivized secretiveness about anything that might prove a liability to one's home department. The net result was that even when suspicions were raised or a complaint made in one place these would not go any further in the epistemic economy of the organization:

> 61. This sense of separation could mean that a concern which arose in one part of the BBC would not be transmitted to or discussed with another part. For example, in 1973, Douglas Muggeridge does not appear to have shared his concern about Savile with anyone in Television. I accept that, if an issue was considered by the Board of Management, it would be known of by senior management across the BBC. For example,

when concerns arose about possible misconduct at *Top of the Pops*, there was some discussion at a meeting of the Board of Management. Soon afterwards, there was discussion about this kind of issue at the Management Director Radio's weekly meetings. But if an issue was not raised at such a meeting, its chances of going across the BBC were slight.

62. At a lower level, there could be a reluctance to discuss a problem which arose in one department with personnel in another. This seems to have been attributable to the sense of competitiveness which prevailed between programme making departments. (Smith 2016).

Now if we consider the BBC as a collective epistemic subject, what we are presented with here is a characterization of an organization whose informational states were radically unintegrated both because of an ethos failing and because of a structural performative failing. The ethos failings consist in the climate of not making complaints about the Talent; combined with the competitiveness between departments. The report also goes into 'the macho culture' in the organization especially around issues of sexual harassment. The structural performative failing is organizational—the fact that different departments were unintegrated and lacked channels of communication between them that would have enabled information sharing:

63. Even within the same programme, there could be difficulties of communication in relation to complaints. Staff working on a programme would not necessarily have the same line manager. Staff working on the production team would be part of a line of management which ran through the producer, maybe to an executive producer and from

there to the head of department. Other staff would have a different management line – for example, the floor manager would not report to the programme producer but to his or her own line manager in the Studio Management Department. That was because the provision of floor management was a central service provided to a programme. Cameramen, sound engineers and audience supervisors had similar separate management structures. This separation seems to me to have the potential for preventing anyone in management from seeing the bigger picture.

The Macho Culture

64. Another reason why complaints or concerns of a sexual nature might not have been passed up the BBC as they should have been related to the 'macho culture' which some witnesses said was present in some (but not all) departments of the BBC. Particular complaint was made about the behaviour and attitudes of technical staff (who were almost entirely male) and of management in Radio 1 and Television's Light Entertainment Department, where there [were] very few women in senior positions. I have the impression that sexual harassment was more common in the Light Entertainment Department and BBC Radio 1 (the areas where Savile worked) than in many other parts of the BBC. Women found it difficult to report sexual harassment. Generally, the attitude of the male managers was thought to be unsympathetic and, of course, there were very few female managers (Smith 2016).

A figure such as Savile could consequently operate relatively freely in his sexual predation in the knowledge that suspicions raised in one department were unlikely to spread, and therefore

unlikely to be treated by any body as evidentially significant. The portrait of the BBC is as an organization which was at that time seriously epistemically unintegrated when it came to the kind of information that was needed to properly pick up on what Savile was doing. People would make complaints, and even if they were believed, the informational content would go nowhere, receiving little to no inferential follow-through. The informational compartmentalization of the organization effectively ensured that the scattered items of information would never amass into a body of evidence, and be treated as such, but would remain inferentially inert epistemic particles dispersed in the organization. The upshot is a portrait of an organization that had serious ethos problems of shielding the Talent, inter-departmental competitiveness and protectionism, and a 'macho culture' especially as regards sexual harassment. In addition to these culpable defects in ethos, there was the significant structural performative failure relating to extreme compartmentalization and consequent inadequate information sharing. All of this adds up, epistemically speaking, to a paradigm example of the institutional vice of inferential inertia. The only upside is that its diagnosis instructively lays bare exactly the kinds of innovations required to improve the situation and to help ensure against recurrences.

I have proposed a conception of epistemic vice such that any culpable lapse in motivational and/or performative elements of epistemic virtue is sufficient for it. And I have applied Gilbert's joint commitment conception of collective agency in order to elaborate what is involved in having an institutional epistemic ethos. An institutional body whose actions systematically betray a jointly committed good epistemic ethos, and/or whose performance systematically fails to implement the good ends of that ethos, is an institution that displays an epistemic vice. Lastly, I have offered a sketch of one institutional epistemic vice in particular, which I have called the

vice of Inferential Inertia. Though it can occur in individuals, its institutional form seems particularly relevant to social life these days. It can occur in both trivial and deadly serious form, and it has been publicly revealed to have found disastrous expression in a valued institution. I hope to have shown it is an institutional epistemic vice worth distinguishing and understanding.[13]

Bibliography

Axtell, Guy (2000) ed. *Knowledge, Belief, and Character* (Lanham Md.: Rowman and Littlefield)

Baehr, Jason (2010) 'Epistemic Malevolence' in H. Battaly ed. *Virtue and Vice: Moral and Epistemic* (Wiley-Blackwell: Chichester, W. Sussex)

Baehr, Jason (2011) *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology* (Oxford: Oxford University Press)

Battaly, Heather (2010) ed. *Virtue and Vice: Moral and Epistemic* (Wiley-Blackwell: Chichester, W. Sussex)

Battaly, Heather (2015) 'A Pluralist Theory of Virtue' in Mark Alfano ed. *Current Controversies in Virtue Theory* (New York: Routledge)

Bondy, Patrick (2010) 'Argumentative Injustice', *Informal Logic: Reasoning and Argumentation in Theory and Practice* 30/3: 263-278

Bratman, Michael E. (1999) *Faces of Intention: Selected Essays on Intention and Agency* (Cambridge: Cambridge University Press)

Bratman, Michael E. (2014) *Shared Agency: A Planning Theory of Acting Together* (Oxford: Oxford Univeristy Press)

Broncano-Berrocal, Fernando and Carter, Adam J. (2020) *The Philosophy of Group Polarization* (London: Routledge)

Cassam, Quassim (2016) 'Vice Epistemology', *Monist* 99 (2): 159-180

Cassam, Quassim (2019) *Vices of the Mind* (Oxford: Oxford University Press)

Crerar, Charlie (2018) 'Motivational Approaches to Intellectual Vice' *Australasian Journal of Philosophy* 96/4: 753-766

Fricker, Miranda (2007) *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press)

Fricker, Miranda (2010) 'Can There Be Institutional Virtues?' *Oxford Studies in Epistemology* (Special Theme: Social Epistemology) eds. Tamar S. Gendler and John Hawthorne: 235-252

Fricker, Miranda (2013) Epistemic Justice as a Condition of Political Freedom' *Synthese* Vol. 190, Issue 7: 1317-1332

Gilbert, Margaret (1987) 'Modelling collective belief', *Synthese* 73 (1): 185-204

Gilbert, Margaret (1989) *On Social Facts* (Princeton NJ: Princeton University Press)

Gilbert, Margaret (2000) *Sociality and Responsibility: New Essays in Plural Subject Theory* (Rowman and Littlefield, Lanham, MD)

Gilbert, Margaret (2002) 'Belief and Acceptance as Features of Groups' *Protosociology* 16: 35-69

Gilbert, Margaret (2004) 'Collective Epistemology' *Episteme* 1(2): 95-107

Gilbert, Margaret (2013) *Joint Commitment: How we make the social world* (Oxford: Oxford University Press)

Konsellman Ziv, Anita (2012) 'Institutional Virtue: How consensus matters' *Philosophical Studies* 161 (1):87-96

Langton, Rae (2018) 'Blocking as Counter-Speech', *New Work on Speech Acts*, ed. Daniel Harris, Daniel Fogal, and Matt Moss (New York: Oxford University Press)

Lewis, David (1979) 'Scorekeeping in a Language Game', *Journal of Philosophical Logic* 8 (1979): 339-359

List, Christian & Pettit, Philip (2011) *Group Agency: The Possibility, Design, and Status of Corporate Agents* (Oxford: Oxford University Press)

Macpherson, Sir William (1999) *The Stephen Lawrence Inquiry: Report of an inquiry by Sir William Macpherson of Cluny* (London: The Stationery Office) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/277111/4262.pdf

Madva, Alex (2019) 'The Inevitability of Aiming for Virtue', *Overcoming Epistemic Injustice* eds. Stacey Goguen and Benjamin R. Sherman (London: Rowman & Littlefield): 85-99

Lahroodi, Reza (2007) 'Collective Epistemic Virtues', *Social Epistemology*, 21: 281-97

Medina, José (2013) *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations* (Oxford: Oxford University Press)

Roberts, Robert and Wood, Jay (2007) *Intellectual Virtues: An Essay in Regulative Epistemology* (Oxford: Clarendon Press)

Sandin, Per (2007) 'Collective Military Virtues', *Journal of Military Ethics* 6(4): 303-314

Smith, Dame Janet (2016) *The Dame Janet Smith Review Report: An Independent Review Into the BBC's Culture and Practices During the Jimmy Savile and Stuart Hall Years* (London: BBC)

Tanesini, Alessandra (2019) 'Epistemic vice and motivation', *Metaphilosophy* 49(3): 350-367

Tuomela, Raimo (2007) *The Philosophy of Sociality* (Oxford: Oxford University Press)

Tuomela, Raimo (2013) *Social Ontology* (New York: Oxford University Press)

Tuomela, Raimo (2017) 'The Limits of Groups, An Author Responds' *Social Epistemology Review and Reply Collective* 6 no. 11: 28-33

Zagzebski, Linda Trinkaus (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press)

[1] See Baehr (2011), Battaly (2010 & 2015), Cassam (2019), Medina (2013). The term 'vice epistemology' was, I believe, coined by Cassam in his 2016 paper of that name.

[2] See Lahroodi (2007); Sandin (2007); Fricker (2010 & 2013); and Konsellmann Ziv (2012). See also, however, the treatment of group epistemic polarization by Broncano-Berrocal and Adam J. Carter (2020), which treats the tendency for a given group to incline towards a more extreme belief than the beliefs of any of its constitituent individual members as a *collective epistemic vice*. I thank Charlie Crerar for directing me to their work on this.

[3] I discuss this case in more detail in Fricker (2013) and my present purpose is to develop more fully some initial ideas I put forward there concerning institutional vice.

[4] For an account of collective agency, specifically those cast as 'we-mode groups', that makes use of a notion of an ethos as part of their characteristic 'we-thinking', see Tuomela (2007, 2013 and 2017).

[5] The germinal opus for the motivationally rich conception of intellectual virtue, which later attracted the label 'responsibilist', is Linda Zagzebski's *Virtues of the Mind* (1996).

[6] Gilbert (1989) is the *locus classicus*; more recently see, for instance, Gilbert (2013); and for a focus on the epistemic see, for example, Gilbert (2004).

[7] In this I may differ slightly from Cassam, who explicitly leaves room for the possibility of vices that are criticisable but non-blameworthy because, for instance, the agents lack the power to correct the intellectual defect in question, such as the cultural prejudice that infected their judgement (Cassam 2019; ix, and ch. 6 esp p. 97). Instead I would tend to categorize such cases as in principle blameworthy—*some* people in the same context after all were able to resist the prejudicial pressures of the day—but where the cultural context of prejudice might make excuses applicable to reduce the appropriate level of blame.

[8] 'Closed-mindedness as a quality of a particular piece of thinking is a *thinking vice*, an epistemically vicious way of thinking or 'thinking style'. It is one thing to *be* closed-minded and another to *think* closed-mindedly' (Cassam 2019; 56).

[9] The example of Galileo's arrogance is from Robert Roberts and Jay Wood (2007; 254), as quoted in Crerar (2018).

[10] See also Tanesini (2019).

[11] Eliana Peck has suggested in conversation that the case of Christine Blasey Ford's testimony to the Senate Judiciary Committee (September 2018) concerning the nomination of Brett Kavanaugh to the Supreme Court is a prime example of this. Blasey Ford was widely considered a manifestly credible witness, and it also seems was in fact *believed*; and yet her testimony may as well have not been believed for all the relevant inferential activity it provoked regarding the question in hand.

[12] Inferential Inertia might also be involved in what Patrick Bondy calls 'argumentative injustice', which is presented as an adaptation of testimonial injustice, since it involves a hearer giving a prejudicially depressed level of credibility to a speaker's argument. (Bondy however focuses exclusively on the effects of negative identity prejudice, rather than prejudice more generally.) See Bondy 2010.

[13] I thank Quassim Cassam and Charlie Crerar for helpful comments on an earlier draft.