

Epistemic agency

Theory and Research in Education

11(2) 135–152

© The Author(s) 2013

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1477878513485173

tre.sagepub.com

Catherine Z Elgin

Harvard University, USA



Abstract

Virtue epistemologists hold that knowledge results from the display of epistemic virtues – open-mindedness, rigor, sensitivity to evidence, and the like. But epistemology cannot rest satisfied with a list of the virtues. What is wanted is a criterion for being an epistemic virtue. An extension of a formulation of Kant’s categorical imperative yields such a criterion. Epistemic agents should think of themselves as, and act as, legislating members of a realm of epistemic ends: they make the rules, devise the methods, and set the standards that bind them. The epistemic virtues are the traits of intellectual character that equip them to do so. Students then not only need to learn the standards, methods, and rules of the various disciplines, they also need to learn to think of themselves as, and how to behave as, legislating members of epistemic realms who are responsible for what they and their fellows believe. This requires teaching them to respect reasons, and to take themselves to be responsible for formulating reasons their peers can respect.

Keywords

Autonomy, epistemic agency, heteronomy, virtue epistemology

The problem of the lists

‘Can virtue be taught?’ Meno asks. When Socrates replies that there is no hope of answering the question until we know what virtue is, Meno obligingly supplies a list: there’s the virtue of a man, the virtue of a woman, the virtue of a child, the virtue of a slave. ‘There is virtue for every action and every age, for every task of ours and every one of us’ (Plato, 1976 [390 BCE]: 72a). Not surprisingly, Socrates demurs. So Meno provides a second list: justice, moderation, wisdom, munificence, and things like that (Plato, 1976 [390 BCE]: 74a). Again Socrates objects. He wants to know what qualifies something for a place on the list. As is well known, Socrates and Meno never arrive at an answer. Neither does Aristotle. Although he links virtue to eudaimonia, it is by understanding how the virtues he cites figure in a good life that we get a sense of what eudaimonia is. The problem persists. Like their predecessors, contemporary virtue theorists, both moral and

Corresponding author:

Catherine Z Elgin, Graduate School of Education, Harvard University, Cambridge, Massachusetts 02138, USA.
Email: catherine_elgin@harvard.edu

epistemological, often rely on lists (Baehr, 2011: 21). No consensus on the criterion for virtue has emerged.

Not that 2,000 years of theorizing has left us completely clueless. We have valuable discussions of the virtues collectively, analyses of individual virtues, and examinations of how the virtues are supposed to figure in a good life. We are evidently in a position to conclude that virtues are relatively stable dispositions to think and/or act well, where a disposition is not merely an ability to do something, but an ability combined with a propensity to do it. Virtues are not mere capacities; they involve a motivational component. Phronesis is a higher-order disposition to adjudicate clashes and to determine where on a continuum a given virtue lies. But poring over lists of (what are said to be) virtues can take us only so far.

The lists are, and acknowledge that they are, incomplete. This prompts some virtue theorists, both ancient and contemporary, to opt for the unity of the virtues. Then rather than, for example, generosity and honesty being different virtues, they are different aspects of (the one) virtue. This simply recasts the difficulty. For it is just as hard to determine which traits are aspects of (the one) virtue as to determine which traits are virtues. Since moving to a unity of the virtues position seems not to help, I will continue to speak of virtues in the plural. I do not think anything of substance hangs on this decision. Whether they are lists of virtues or lists of aspects of virtue, they end with 'etc.' or 'and the like'. This would be unobjectionable if a way forward were clear. But the items on any given list tend to be disparate: honesty, justice, magnanimity, and the like; open-mindedness, rigor, curiosity, and so forth. The items listed are so motley that we are at a loss to tell what else to include. Some items on familiar lists are controversial. Contrary to Aristotle, Amélie Rorty argues that courage is not a virtue (Rorty, 1986). Many contemporary readers harbor doubts about the status of magnificence. Whether piety and patriotism are virtues is doubtful. To make matters worse, different, individually plausible lists are mutually inconsistent. Aristotle considers proper pride a virtue, humility a vice. Humility, he thinks, consists in underrating yourself, failing to give yourself your due. Christians consider humility a virtue, and any degree of pride a vice. How should this be adjudicated? The assurance that the phronimoi know the answer is unhelpful, without a non-question-begging way to identify the phronimoi.

One might be tempted to dismiss Socrates' demand as important only to Platonic essence-mongers. Perhaps the rest of us should settle for a list, even if it ends with 'and so forth' (possibly muttering sotto voce that 'virtue', like 'game', is a family resemblance concept). Although I am about as far from a Platonic essence-monger as one can get, I am not convinced that virtue theorists can afford to be quite so cavalier. If we want to foster virtue in ourselves or our students, we need to know whether a given trait is a virtue. Should we encourage or chastise the child who takes pride in her own accomplishments? Should we promote compromise and consensus building in the classroom or call it giving in to peer pressure? Or, as seems more likely, do we need to ask more nuanced questions to decide which behaviors are virtuous? However we settle such matters, we need to know what makes virtues valuable. It is not enough to say with Aristotle that the virtues contribute to human flourishing. For their contributions are not just causal; the virtues are in part constitutive of flourishing. We cannot independently characterize flourishing, then dissect it to identify the virtues.

The problem of the lists afflicts both ethical and epistemic virtue theories. Here, I will focus on epistemic virtue theories (although I believe that some traits that are typically considered ethical virtues are required for epistemic virtuousness as well (Elgin, 2011).) My goal in this article is not to argue for or against epistemic virtue theory, but to suggest a way that epistemic virtue theory can get beyond listing purported virtues and appealing to intuitions or common sense for consensus that particular items on the list indeed are virtues.

Virtue epistemology

Virtue epistemologists maintain that the epistemic status of an agent underwrites the epistemic status of her beliefs. Rather than the agent's epistemic excellence deriving from her harboring epistemically estimable beliefs, the epistemic status of those beliefs derives from the excellence of her epistemic character. A true belief qualifies as knowledge, they contend, only if, in forming and sustaining that belief, the agent is epistemically virtuous – only if, that is, in forming and sustaining the belief, she exhibits stable dispositions to be conscientious, scrupulous, open-minded, rigorous, and so forth. Whatever its other merits, a true belief fails to qualify as knowledge if it does not stem from epistemically virtuous dispositions.

It is useful to divide virtue epistemology into two camps – virtue reliabilism and virtue responsibilism. (There are also hybrid positions but for current purposes we can neglect them (Axtell, 1997).) Although contemporary epistemology contains responsibilists and reliabilists of other stripes, here I focus on virtue reliabilism and virtue responsibilism.

According to *virtue reliabilism*, what makes a disposition an epistemic virtue is that it is reliably truth conducive. Its exercise need not, of course, yield truth in every case. But its exercise should be far more likely to yield a truth than to yield a falsehood and be far more likely to yield a truth than its non-exercise does. Attentiveness to evidence is a reliabilist virtue then just in case beliefs formed while being attentive to evidence are in general far more likely to be true than beliefs that are arrived at by ignoring or slighting evidence. According to *virtue responsibilism*, what makes a disposition an epistemic virtue is that exercises of that disposition generate or sustain responsibly held beliefs. Attentiveness to evidence is a responsibilist virtue just in case it would be epistemically irresponsible to ignore or slight the evidence when forming a belief about an empirical matter.

Virtue reliabilists have the resources to answer Socrates' question. Truth, they maintain, is our overarching cognitive objective. Since what makes a disposition an epistemic virtue is that it is truth conducive, the list of epistemic virtues includes all and only those dispositions that in fact reliably promote the generation of true beliefs and inhibit the generation of false beliefs. For example, whether open-mindedness or curiosity is an epistemic virtue depends on whether a belief that emerges from a disposition to open-mindedness or curiosity is far more likely to be true than a belief generated randomly or by an antithetical disposition. Acceptable epistemic standards screen for reliability.

This is an attractive picture, but worries remain. Although it is plausible that dispositions we typically consider epistemic virtues are truth conducive, the question is an

empirical one and the answer is not obvious. Intellectual courage – the disposition to hold fast to a belief despite opposition, so long as you consider your reasons sound – is typically considered an epistemic virtue. But whether the beliefs to which epistemic agents hold firm are far more likely to be true than beliefs that epistemic agents readily sacrifice is an open question. The answer probably depends on other traits of the believers. Suppose, for example, that agents who display intellectual courage turn out to be rather bad at assessing evidence. Then their steadfastness might lead them to hold fast to beliefs they ought to abandon, since despite their considering their reasons sound, the evidence tells against them. Or, to take an extreme case, consider religious martyrs who willingly die rather than abandon their beliefs. Pretty clearly they consider their reasons sound; otherwise, when faced with death (typically a painful death), they would recant. Such martyrs display intellectual courage. But they are not all holding fast to true beliefs. Indeed, given the diversity of conflicting creeds for which people have been martyred, it is likely that many more have been martyred for false beliefs than for true ones. Perhaps the number of acts of holding fast to true beliefs under less harrowing circumstances swamps the number of martyrs for false religious beliefs. If so, the reliabilist could still maintain that intellectual courage is reliably (albeit not universally) truth conducive. But we do not know. The problem generalizes. It is an open, empirical question whether the various dispositions that we take to be epistemic virtues actually are truth conducive. In any case, few if any epistemic virtues are truth conducive on their own. At best, a cluster of virtues, abilities, and background beliefs in an epistemically favorable environment is truth conducive. Virtue reliabilism must specify what traits belong to the cluster and determine whether clusters of different dispositions might be equally truth conducive. With respect to any candidate virtue or cluster, the position is hostage to epistemic fortune – to its being the case that that disposition or cluster of dispositions is actually truth conducive.

Truth-conduciveness is a problematic criterion for epistemic virtue in any case. It is hard to believe that epistemic virtue is exhibited by people who relentlessly pursue trivial truths and reliably achieve them. Take any number, and keep adding two. Each sum reliably yields another truth. Or take any truth, and add a disjunct. You now reliably have another truth. Although these methods of generating true beliefs are far more reliable than, for example, properly amassing and judiciously assessing evidence, they seem clearly deficient in epistemic virtue. Perhaps we could devise a criterion for reliably generating truths worth having, but how to do so is far from obvious.

An attractive feature of epistemic virtue theory is that it can both acknowledge that, and explain why, virtuous epistemic agents are cognitively admirable even in epistemically hostile environments. Regardless of circumstances, the scrupulous, data gathering, reason assessing scientist seems epistemically more admirable than the cavalier, intellectually fickle jumper to conclusions. This is so even in a demon world where neither is reliable, indeed even in a demon world where the jumper to conclusions is lucky enough to be right more often than the scrupulous scientist. Despite the hostile epistemic environment, one wants to say, the scientist is virtuous in making the best possible use of her epistemic resources; the jumper to conclusions is not. Virtue reliabilism has to deny this. If neither is reliable in a demon world, neither is virtuous; if the jumper to conclusions is more reliable, she is more virtuous. So although reliabilism can underwrite its list of

virtues, I am not optimistic about its prospects. Truth-conduciveness seems neither necessary nor sufficient for epistemic virtue.

Does virtue responsibilism fare any better? It maintains that virtuous epistemic agents are epistemically responsible agents; they form, sustain, and revise their beliefs, methods, and standards as they should. Their epistemic virtues are the dispositions that simultaneously equip and motivate them to do so. Virtue responsibilism has the merit of being able to handle skeptical, and semi-skeptical scenarios.¹ Making the best use of available resources is what a responsible epistemic agent should do, whether the environment is hostile or auspicious or something in between. We can typically tell whether someone is disposed to act responsibly, even though we do not know whether the dispositions that give rise to his actions are reliably truth conducive. Open-mindedness and sensitivity to relevant detail qualify as responsibilist virtues, though we lack the empirical evidence to determine whether they tend to generate a preponderance of true beliefs. Finally, virtue responsibilism does not get tripped up by triviality. The responsibilist can say that, even if a disposition to robotically generate beliefs of a particular sort is truth-conducive, those are not truths worth having, so it is irresponsible to foster or exercise that disposition.

Still there are problems. One is that my characterization of virtue responsibilism is practically platitudinous. *Of course* virtuous agents generally do as they should, and to be virtuous is to generally do as one should; *of course* vicious agents frequently fail to do as they should, and to be vicious is to frequently or flagrantly fail to do as one should. Who didn't know that? The worry is not that my characterization of virtue responsibilism is in terms of other normative notions. There is no requirement that epistemic virtues reduce to or be definable in terms of something non-normative. Rather, the worry is that the characterization seems not only to be circular, but to delimit a circle with an exceedingly small diameter. To say that responsibility consists in doing what virtue requires, and virtue consists in acting responsibly gets us nowhere. The question that hasn't been answered is what determines what an epistemic agent should do? Why should we believe that acting as the virtues on our list prescribe is responsible? In short, why should we think and act as responsibilist virtue theory requires?

Our predicament is this: according to virtue responsibilism, beliefs formed and sustained on the basis of virtuous dispositions are responsible (or responsibly held) beliefs. To call open-mindedness a virtue then is to say that an agent should seriously entertain a reasonable range of alternatives before forming a belief. To call attentiveness to evidence a virtue is to say that an agent should carefully amass and judiciously assess evidence before forming a belief. And so on. We can run down our favorite list, articulate the epistemic recommendations they encode, and say it is responsible to follow those recommendations. Then we can run down our favorite list of vices – credulousness, fickleness, dogmatism, and so on – and explain why they are epistemically irresponsible. But we are still stuck with a list.

Nor is this the only difficulty. Like virtue reliabilism, virtue responsibilism confronts the problem of extending and emending the list. And few, if any, epistemic virtues give rise to responsibly held beliefs on their own. Rather, the exercise of a cluster of epistemic virtues is required to generate a responsibly held belief. So, like virtue reliabilism, virtue responsibilism has to characterize the cluster and has to determine whether there is a

uniquely optimal cluster. Moreover, responsibility involves obligation. If it is your responsibility to return the books to the library, you are under an obligation to do so. But even if a disposition is epistemically good, it does not follow that agents have a responsibility to develop and exercise it. There are plenty of dispositions that it would be good to have and to exercise which do not place us under any obligation. (It would be good to develop the disposition to clean your closet more frequently, but you are under no obligation to do so.)

One familiar requirement on responsibility is 'ought' implies 'can': x is responsible for y , only if whether y obtains is under x 's (direct or indirect) control. If x cannot bring about y , or x cannot prevent y , then x is not responsible for the occurrence or non-occurrence of y .² Epistemic agents have little if any control over their epistemic environment. They cannot, for example, control whether they live in a demon world. So, the responsibilist would say, they are not epistemically remiss if the falsity of their beliefs derives from the undetectable machinations of a malevolent demon. Nor are agents epistemically remiss if, in a more auspicious world, they form a false belief on the basis of evidence that is scrupulously garnered and carefully analyzed, but misleading. Truth conduciveness then is not the criterion of epistemic responsibility. What is the alternative? Even in an inauspicious world, agents can exercise epistemic self-control. They can decide whether to endorse considerations that present themselves as candidates for belief.

Reflective endorsement

According to Kant (1981 [1785]), an autonomous agent makes the laws that bind her; a heteronomous subject is bound by constraints that he neither makes nor endorses. The latter is not strictly an agent, for he does not act; he merely reacts. Kant focuses on behavior, not people, as autonomous or heteronomous. Real people (being neither holy wills nor robots), are capable of, and presumably engage in, both sorts of behavior. In speaking of subjects as one or the other, I employ a simplified model that directs attention to the extremes of a motivational continuum. There is no suggestion that the person I describe as heteronomous cannot behave autonomously, or that the one I call autonomous cannot behave heteronomously. Nor is there any suggestion that any actual person could be entirely autonomous or entirely heteronomous. Indeed, it is because people have the capacity for both autonomous and heteronomous behavior that we can be held responsible for what we do.³ Although Kant describes the heteronomous subject as driven by his inclinations or desires, action is a joint product of belief and desire. So an agent is equally heteronomous if he is driven by beliefs that are not under his control. This suggests that it may be useful to import Kant's distinction into the epistemic realm.

Inasmuch as we are cognitive beings with sensory and representational capacities, a variety of things simply strike us as being the case. Deliverances of sensation, emotion, apparent memory, and imagination present themselves as candidates for belief. It might seem that this makes us, at base, epistemically heteronomous. We can't help but see, hear, and feel what we do. But the issue is not whether we are passive with respect to inputs; rather, it is what we do with them. The heteronomous subject simply accepts deliverances at face value. For him, seeing is believing. Or perhaps seeing is not always believing, but whether he believes what he sees is not up to him. He simply finds himself

believing something on the basis of his deliverances, being under the sway of whatever belief-forming mechanisms happen to be operative. Since his beliefs are considerations that just strike him, since he can neither defend nor criticize them, they are not under his control.

The beliefs he finds himself with might be reliable. Perhaps humans evolved in such a way that what strikes us as dangerous usually is dangerous. If so, when he thinks the snake is dangerous it usually is; when he thinks the spider is harmless it usually is. Maybe this generalizes. If the subject's dispositions are reliable, he is objectively secure. Nevertheless, he is in a subjectively precarious position. Reliabilism holds that a disposition or process is epistemically creditable just in case it is reliably truth conducive; a true belief is knowledge just in case it is a result of exercising a reliable disposition. Whether this is so is independent of anything the subject thinks about it. As far as reliabilism is concerned, heteronomous but reliable beliefs, processes and dispositions are epistemically unobjectionable. Still, since the subject cannot reflectively endorse or reflectively repudiate his beliefs or the dispositions or processes that give rise to them, he is a victim of circumstances.

One might argue that the situation is not so dire as I make it out to be. The believer can attend to his track record. He can register his successes and failures, and credit the dispositions and mechanisms that give rise to success. This is surely right. But the question then arises for the second-order beliefs, processes and dispositions he credits. Do they just strike him as correct? Can he defend the way he identifies, records, and tallies successes and failures? Even if his methods are reliable, so long as he has nothing more to say than that it strikes him that this is a good way to proceed, to identify, to tally, or to measure, he remains vulnerable. Second-order heteronomy is as problematic as first-order heteronomy.⁴ Moving to yet higher orders is of no help. So long as the best that can be said for the principles or processes he uses is that they strike him as correct, he is vulnerable. He may be on safe ground, but he has no reason to think so. He is being driven to believe by considerations he happens to be struck by. And he is not responsible for what strikes him.

Is the autonomous agent any better off? She too may be struck by things. Her inputs may be exactly the same as those of the heteronomous subject.⁵ But her relation to them is different. Forming beliefs is something she does, not something that happens to her. What she does in forming beliefs is, as it were, filter deliverances (which may be putatively perceptual inputs, apparent memories, or simply ideas that pass through her mind) through a critical sieve, accepting only those she considers worthy of her reflective endorsement. To see the difference, consider someone blind from birth.⁶ Not wanting her to feel handicapped, her parents, teachers and peers never let on that others have a mode of sensory perception that she lacks. Hence she has no idea that objects are visible. Then she has eye surgery. When the bandages are removed, she is struck by a rich and complex array of visual inputs. Initially, I believe, she would have no reason to think that they are sensations *of* anything rather than like the veils, floaters and flashes characteristic of retinal detachment. She certainly has, at the outset, no reason to consider them reliable. Only after she has discerned regularities in the sensory flux, correlated visual sensation with the tactile sensations she already considers reliable, and perhaps spoken with others about her visual experiences does she have any reason to credit her visual sensations. She

needs not only to have the sensations, but to be in a position to reflectively endorse their deliverances.

To be sure, an epistemic agent does not entertain deliverances one by one and ask himself, 'Am I buying this?' Rather he develops and deploys a variety of methods, mechanisms, heuristics and habits that enable him to credit or discredit wide swaths of inputs efficiently. Apparent sightings of not unexpected middle-sized objects in good light in the center of his visual field are apt to be readily accepted. Having no reason to reject them may constitute reason enough to accept them. But if there is something incongruous about a deliverance – if, for example, he seems to see a camel on Main Street in small town America – he does not simply accept it. He may reject it as a hallucination, an illusion, or a misperception; or he may suspend judgment until he investigates whether a camel is part of a local publicity stunt, political rally or a circus parade. Likewise, if there is something untoward about his condition as an observer. If he is half awake or on new medication or intently focusing on something else, he may refuse to credit a surprising deliverance. The critical point is that his beliefs are products of his reflective endorsement. He is willing to stand behind them because they satisfy his standards. What makes the epistemic agent responsible for his beliefs is that he takes responsibility for them.

Reflection does not occur in a vacuum. It involves sensitivity to epistemic ends and means. It is imbued with (often tacit) background assumptions and responsive to epistemic circumstances. A sighting of what looks to be a camel on Main Street warrants further investigation; a sighting of what looks to be a car on Main Street or a camel in Giza typically does not. An agent's reflective endorsement is a willingness to be bound by an epistemic commitment (a belief, method, or standard) because she thinks that her epistemic purposes will be served by her being so bound. She considers the items she reflectively endorses trustworthy bases for inference and action when her ends are cognitive.

It might seem that this too leaves the agent vulnerable. She is willing to stand behind the commitments she reflectively endorses. But should she be? The reliable but heteronomous subject was vulnerable because he could not reflectively endorse his beliefs. It might seem that the autonomous agent, as I have characterized her, has the opposite problem. She can reflectively endorse her beliefs, but they might not be reliable. She is perhaps subjectively secure, but remains objectively at risk. In fact, if her personal standards are sufficiently skewed, her reflective endorsement may be epistemically worthless. Can we evade an untenable subjectivism here?

When an agent's beliefs are ones she reflectively endorses, they are at least *prima facie* defensible. She can give what she takes to be reasons for them. Philosophers distinguish between internal and external reasons (Williams, 1981). An external reason is a consideration that actually supports a claim. Some external reasons are known to no one: 'There is a reason why fewer children have elevated cholesterol levels even though childhood obesity rates remain constant' is true, even though no one currently knows that that reason is. Others are known to some people, but not to the agent harboring the belief. There is a reason to believe that Deb will get the part, and that reason is known to the casting committee. But, although Fred believes she will get the part, he does not have any reason to believe it, as he has never seen her act, is not privy to the deliberations of the casting committee, and the decision has not been announced. Assuming an agent is not

remiss in being ignorant of them, external reasons do not bear on her epistemic responsibility. The only reasons a person can appeal to are internal reasons – reasons within her epistemic purview. To say that an agent's belief that chicken soup alleviates colds is defensible is to say that she can defend it. Her reason is something that she can adduce to support her claim.

The limits on internal reasons are not entirely clear. Among the considerations within the agent's epistemic purview are those she can call to mind. She has learned them and has not forgotten them. Let us call these her available reasons. If she recognizes their relevance, she can surely appeal to them to defend her claim. But arguably her purview is wider than that: it includes information she can readily draw on – information obtainable in her epistemic milieu, whether she is currently privy to it or not. We might call such reasons accessible reasons. Sometimes we take available reasons to set the limits on accountability: 'Phil had reason to believe that Jane is old enough to vote, since he knew full well that Jane is older than Jim and that Jim can vote.' Sometimes we appeal to accessible reasons: 'Meg had reason to believe that the course had a prerequisite, since it was clearly stated in the catalog.' Accessibility and availability are matters of degree. Some reasons are readily accessible; others take considerable digging. Some available reasons are easily called to mind; others need to be dredged up. Insofar as epistemic responsibility involves responsiveness to reasons within the agent's ken, just how available or accessible those reasons must be is a question that needs to be addressed. It is not one I will address here.

It is not enough that the agent has something she considers a reason. If her claim that chicken soup alleviates colds were challenged, she could not prevail by saying that she read graffiti to that effect. Graffiti is notoriously untrustworthy, so her interlocutor would and should dismiss her appeal to graffiti as no reason at all. Reasons, as Korsgaard (1996) argues, are considerations we give to each other. They are not mere expressions of personal conviction, but considerations our interlocutors should countenance. The agent could give as her reason that she had amassed and evaluated the relevant statistics, that she read about the correlation in *The New England Journal of Medicine* or *The New York Times*, that she had been told by her physician that this particular old wives tale is true. She might even say that her grandmother told her. Assuming that her grandmother is not an expert on contagious diseases, this would be a relatively weak reason. But if the members of her epistemic community were inclined to give it some epistemic weight, it would still be a reason.

Even if no one ever asks, she has a reason if she is prepared to adduce such a source if called upon. But because she could not responsibly recommend that others share a belief if her only basis for it was something she read on the bathroom wall, the graffiti is not a reason for her. If she thought that graffiti was a trustworthy source of information, she would be mistaken. Reasons – even internal reasons – have to stand up to intersubjective scrutiny.

A consideration's status as a reason is not a function of the mindset of its audience. The statistical evidence the agent amassed is a reason even if her audience is ignorant of statistics; and it is a reason even if she never reports on her findings. The issue is not what considerations a given audience does accept as reasons, it is what considerations they should accept. What determines that?

Here it pays to delve more deeply into Kant. One formulation of the Categorical Imperative is that an agent should act only on a maxim that he could advocate and accept as a legislating member of a realm of ends. Extending that idea to epistemology, an epistemic agent should believe only considerations that she could advocate and accept as a legislating member of a realm of epistemic ends. Let us call this the epistemic imperative.

Agents in the realm of ends make the laws that bind them. They are not merely subject to the laws that govern their behavior, they are enactors of those laws. Since they would not make laws that they considered it inappropriate to be bound by, they think it right that they be subject to the laws of the realm of ends. They reflectively endorse those laws. Similarly, I suggest, for epistemic commitments. In making and reflectively endorsing commitments, the agent exercises her autonomy. She considers herself justifiably bound by those commitments because she believes that being bound by those commitments will promote her epistemic ends, given her epistemic resources. The ends may, as the reliabilist maintains, be truths; they may be truths that satisfy further standards (for example, standards of significance or relevance); they may be non-truths (perhaps non-propositional representations, effective models, informative idealizations or illuminating fictions) that she takes to yield information or insight about the topic that concerns her. The critical point is that agents set their ends. So, even if an agent agrees with the reliabilist that the overarching end is truth, her relation to that end is different from his.

It might seem that reflective endorsement yields only subjective value. The agent sets epistemic ends for herself, makes epistemic commitments for herself, and on reflection is willing to stand behind them. This assures a measure of internal consistency, but not much more. But Kant does not say that a maxim is acceptable only if it is something an agent could enact as the philosopher king of a realm of ends. Satisfying merely personal standards is not enough. It is important that the word 'legislators' is plural. Legislators work together to enact laws. To be effective, they must convince their colleagues of the acceptability of the legislation they propose. This requires that the basis for their recommendation be publicly articulable or displayable⁷ and justifiable to other legislators in light of their shared commitments and goals. The realm of ends is actually a commonwealth. The same holds of epistemic commitments. The reasons for them must be specifiable and justifiable to the other members of the epistemic community. Since the realm of epistemic ends is supposed to be the arena within which epistemic agents live their cognitive lives, those commitments must mesh. Not only must each be individually acceptable; all must be collectively acceptable. This means that there are consistency and coherence constraints on what the legislators can endorse.⁸

Commitments bind: they constrain what it is permissible to do and how it is permissible to do it. Recognizing that there is nothing epistemically special about her, an individual epistemic agent does not think that she should be bound by commitments that other similarly situated epistemic agents need not be bound by; nor should she be free of commitments that bind other such agents. She does not, for example, consider her commitment to eschew the gambler's fallacy a personal predilection. She does not think that, although she prefers not to commit the gambler's fallacy, other equally adept agents are within their epistemic rights to decide otherwise. She considers the gambler's fallacy a

mistake: everyone should avoid committing it. But she recognizes that epistemic agents should be bound only by commitments that they can reflectively endorse. So she takes it as a criterion on the commitments that she can reflectively endorse that similarly situated epistemic agents should be able to reflectively endorse them as well.

The epistemic imperative directly vindicates many familiar epistemic virtues. Consider open-mindedness: an epistemic agent cannot reflectively endorse a contention unless she is in a position to believe that other members of the epistemic community could endorse it as well. And she is in no position to responsibly believe that without entertaining the alternatives that they might consider credible. So she must act open-mindedly in developing the contentions she advocates. Her open-mindedness is not boundless: it need not extend to examining hypotheses involving space aliens, since those hypotheses are not ones that her intellectual compatriots would, or by their own lights should, take seriously. Rigor, responsiveness to evidence, and impartiality receive similar treatment. In each case, the scope of the demand that a virtue places on the agent is delimited by the epistemic community. Each agent stands in a reciprocal relation to other members of the community. Besides venturing only hypotheses that she thinks worthy of their reflective endorsement, she must be in a position to responsibly reflectively endorse or reflectively repudiate their hypotheses. She needs to be able to think and act impartially, fair-mindedly, and knowledgeably.

Besides the virtues that are directly vindicated by the epistemic imperative, others are vindicated by what we have come to understand about our epistemic situation. As we learn more about a topic, we learn more about how to learn about that topic. We devise new methods and design new instruments. We discover hitherto unforeseen limitations or biases that cast doubt on results generated by old methods and instruments. We uncover mistakes and diagnose their sources. Considerations that our predecessors might have reflectively endorsed no longer warrant reflective endorsement. Considerations they might reasonably have refused to endorse, we are in a position to endorse. Two hundred years ago, physicists would readily have endorsed the contention that space is Euclidean, and would have adamantly refused to endorse the contention that mass increases with acceleration. Now we know better. Because the history of inquiry shows that advances in understanding often involve correcting the errors of our forebears, fallibilism is a virtue, dogmatism a vice. An epistemic agent, seeing that other agents who satisfied the standards of their communities of inquiry turned out to be wrong, should recognize that even though she satisfies the standards of her community she might be wrong as well. Intellectual humility is also called for. She cannot claim that her contentions are permanently acceptable, only that they are as reasonable as any available alternative in the current epistemic circumstances. In proffering them to the epistemic community, that is what she is willing to defend.

I will not run through the whole list of epistemic virtues. The examples I have sketched indicate how the epistemic imperative underwrites the virtues and why responsibilism considers behaviors that accord with the virtues obligatory rather than merely a good thing to do (like cleaning your closet more often). One cannot function as a legislating member of a realm of epistemic ends without performing the sorts of acts that fall under the virtues.

Division of labor

A critical question concerns the constitution of the community of epistemic ends. Who belongs to it? More to the point, what qualifies someone for membership? Kant might say that the community of epistemic ends consists of all rational agents throughout history. The problem is that little is separately justifiable to, or reflectively endorsable by, all rational agents. What an epistemic agent is in a position to reflectively endorse depends heavily on her cognitive background. Much that the community of particle physicists reflectively endorses is unintelligible to the rest of us. But it would be wrong to dismiss their epistemic achievements because of our ignorance or incompetence.

The resolution lies in the division of epistemic labor. Very roughly, a community of inquiry consists of those who share background beliefs, methods, standards, and goals. The way that they share is critical. Not all members of the community have to have the same background beliefs. The particle physics community includes experimentalists, theoreticians, and statisticians, each of whom relies on the expertise of the others. That community overlaps to a considerable degree with other communities of physicists – plasma physicists, computational physicists, atomic physicists, etc. – and with other scientific communities – chemistry, astronomy, mathematics, and so on. All overlap with and rely on communities of experts in the design, calibration, construction, and programming of their instruments. Contentions about particle physics should respect the standards set and reflectively endorsed by the particle physics community, many of which are shared with or informed by those of overlapping communities. If they respect those standards, they are epistemically responsible. To respect the standards of a community is to be responsive to them. This does not require satisfying those standards. Rather, it requires either satisfying them or having and being able to articulate a cogent reason to revise or reject them.

Earlier I said that responsibilism does not require vetting each candidate for epistemic acceptance individually. We routinely endorse deliverances from creditable sources – memory, perception, and so on. One such source is testimony. On her own, each of us is in a position to know very little. To compensate for our individual limitations, we rely on the testimony of others. Just as there is a division of cognitive labor within the particle physics community, there is a division of cognitive labor across the realm of epistemic ends. Just as experimental physicists defer when appropriate to theoreticians, amateurs defer to physicists. The relevant communities need not be scientific. We defer to the trained auto mechanic for an explanation of the ominous rattle in the engine, to the knowledgeable historian for an understanding of the influence of the cotton gin on American slavery, to the experienced parent for insights into how to soothe a teething baby.

Still, there are communities and communities. We might be fairly sanguine about saying that satisfying community standards suffices for epistemic responsibility in particle physics or auto mechanics. But what about the community of psychics? Are we forced to say that if they have managed to contrive, reflectively endorse, and by their own lights satisfy standards for contentions about the paranormal, we should credit their claims? We are not. For we are in a position to assess their standards, methods, and claims.

One measure of epistemic acceptability is predictive success. Not every discipline makes predictions; but those that do are subject to assessment on the basis of their track records. For a predictor to be epistemically creditable, its predictions should be borne out considerably more often than chance, considerably more often than what one would expect if its distinctive causal commitments were dropped, and considerably more often than the predictions of its rivals. The predictions must be definite enough that it is possible to tell whether they are borne out. And it must be determinate whether their being borne out redounds to the credit of the putative predictor. Psychics make predictions. If the standards of their community are satisfied by predictions that are too vague to be tested, are not borne out when tested, or simply replicate what would otherwise be expected, then satisfying their standards does not make for epistemic credibility. When Professor Trelawny, on reading his tea leaves, announces that Harry Potter has a mortal enemy, Hermione replies, 'Everybody knows that!' (Rowling, 1999: 106). Coming up with that insight is no reason to credit divination. If we are to trust divination, Hermione rightly intimates, it should reveal things that we would not otherwise have known. Similar points can be made about explanatory success, systematicity, fit with the commitments of neighboring areas of inquiry, and so forth. The familiar list of theoretical and empirical desiderata can be used to assess the claims of a putatively cognitive enterprise.

My point here is not primarily to criticize psychics. It is to emphasize that there are generic demands that creditable communities of inquiry must meet. Not only must the commitments be internally coherent and consistent, they must cohere with and be consistent with other things we have reason to believe. If the community inquires into empirical matters, its claims need to be backed by evidence. If it makes predictions, the predictions must (often enough) be borne out. These are entirely familiar requirements. Communities of inquiry make fine-grained commitments that, given their understanding of their topic and the effective ways of investigating it, are locally appropriate realizations of more generic coarse-grained commitments. The failure of their fine-grained commitments to satisfy (or at least approximate) coarse-grained requirements, unless backed by strong reasons to think that the fine-grained commitments need not satisfy them, is a reason to refuse to reflectively endorse their findings.

This may seem too weak to evade pernicious relativism. Why shouldn't we discredit the standards of the community of psychics by recognizing that they violate a priori standards to which all communities of inquiry are subject? One reason I am reluctant to make this move stems from skepticism about the a priori. But there is another reason, which should be compelling even to those who do not share my Quinean scruples. Since 'ought' implies 'can', a subject is under no obligation to do anything she cannot do. In the epistemic realm, I suggest, what an agent *can* do is circumscribed by the methods and reasons accessible to her. But principles – even a priori principles, if there are any – are discovered. Prior to their discovery, they were inaccessible. So agents were not then remiss for failing to draw on them. Since, for example, the predicate calculus was inaccessible to Leibniz, his failure to use it or appeal to it in his reasoning was not epistemically irresponsible. In one sense, of course, if there are a priori principles, anyone can follow them. But this is a sense that makes them external reasons, not reasons an arbitrary agent, anywhere, at any time could appeal to.

Since 'ought' implies 'can', the standards the responsible epistemic agent must respect are not those of some idealized, timeless community of rational agents. She could have no idea what the standards of such a community are or what it would take to satisfy them. Rather, she must respect the current or reasonably foreseeable standards of the realm of epistemic ends she inhabits. Just what these standards are is not determined a priori. They emerge over the course of inquiry, as the community learns about the subject matter, how to investigate the subject matter, and what sort of understanding of the subject matter is practically possible in the epistemic circumstances, given the resources at hand, and the questions the community wants to answer.

Nevertheless, it might seem that the extent to which we have to rely on one another undercuts any claim to epistemic autonomy. Often, it seems, we amateurs are in no position to defend our own beliefs, as they depend on fine-grained epistemic commitments that we are not privy to and would not understand even if we were. Physics is an empirical science. So we know that its contentions are answerable to evidence. We therefore know that we ought not to accept the contention that the Higgs boson exists unless there is sufficient evidence. But we defer to the community of particle physicists to specify what counts as evidence, what counts as sufficient evidence, and when such evidence is in hand. We have no choice. Still, the responsibilist insists that epistemic commitments must be defensible. If my defense of the contention that the Higgs boson exists consists in saying, 'They said so', it looks weak.

Actually it is not. Although expertise is distributed across the epistemic community, this does not undermine the responsibility of the individual epistemic agent. She confers epistemic authority on those she counts as experts; and she retains the right to revoke it. She can and at some point should decide that the auto mechanic does not know what he is talking about (even though she does not know what a catalytic converter does), that the investment counselor's advice is not credible (even though she is only dimly aware of what a hedge fund is), that the physician who links autism with vaccines is untrustworthy (even though she has no idea why the incidence of autism is rising).

The requirement that autonomous epistemic agents think for themselves is not in tension with the recognition of the extent of our epistemic interdependence. Rather, it means that each agent is ultimately responsible for her choice of experts. There are no doubt chains or, more realistically, webs here: the patient defers to her physician who in turn defers to networks of medical researchers, who in turn defer to networks of biochemists. Nor is the deference primarily to a person: the expert is taken to embody the epistemic commitments of his area of expertise. If a contention's defense consists of 'They say so' backed by defensible reasons for thinking that *their* saying so is trustworthy, it is not a weak defense; it is an exercise of epistemic authority.

Educating for autonomy

This has clear implications for education. We need to educate students for epistemic autonomy – to equip them with the knowledge, abilities, and motivations needed to function as legislating members of a realm of epistemic ends. Students should learn to think of themselves as epistemic agents – that is, as personally responsible for what they believe. That means that they need to learn how to think for themselves. They

need to learn how to *take responsibility* for what they believe, and why it is important to be able to do so. They should not only deploy methods that are in fact trustworthy, but should also understand that, why, and to what extent they are trustworthy. Students should learn not just that history teaches that p and that science tells us that q , but also how science and history arrive at their conclusions, and both why and to what extent conclusions arrived at in this way are credible. They should learn when to defer to experts, and what qualifies someone as an expert. And they should learn when to be skeptical of what the experts say. That requires critical reasoning skills. Students should learn to believe only what they can reflectively endorse, and reflectively endorse only what they can defend. They should recognize that to defend a belief requires backing it by reasons other members of the epistemic community can reflectively endorse as well. They should learn to exhibit epistemically virtuous behavior not by rote (give three reasons for your hypothesis, three reasons against it, assess those reasons and draw a conclusion), but because they appreciate how epistemically virtuous behavior contributes to or figures in developing and sustaining defensible beliefs. That is, they should understand why it is a good idea to be open-minded, rigorous, intellectually honest and the rest. And they should appreciate that the methods, standards, and results they are learning are revisable.

To an alarming degree, contemporary education fosters heteronomy. Facts, methods, and practices are presented as things students simply have to master. Students are not encouraged to think about why or whether these facts, methods and practices contribute to or figure in an understanding worth having. They are told, for example, that in writing a term paper they must cite their sources. But they are not invited to reflect on how dependence on sources evinces the powers and limitations of the division of epistemic labor. They are told to put things in their own words. But no one explains why doing so is worthwhile. If they recognized that in writing an essay they are taking epistemic responsibility for what they say, they would appreciate the desirability of saying things in a way that they can defend, rather than just parroting back something they've read. Students are told to perform an experiment in the laboratory. But they are not asked to reflect on the nature, powers and limitations of experimental design, on how what they find in the laboratory might differ from what goes on in the wild, on why different experimenters might arrive at different results, or what to make of it if they do. Nor is it merely a matter of asking them to reflect on such matters. They need to be taught how to do so – that is, they need to be taught what sorts of considerations are relevant in thinking about such matters, and how to weight those factors.

I am not claiming that current demands on students are presented as arbitrary fiats. On the contrary, reasons are often given. But those reasons tend to be epistemologically exogenous – to pass the test, to get out of 10th grade, to avoid being charged with cheating. Such reasons might well be good motivators. But, being heteronomous, they yield only hypothetical imperatives. If you want to pass trigonometry, you should learn the law of cosines. If, however, you do not care about passing, or despair of passing anyway, such a reason has no claim on you. Moreover, such a reason gives you no particular purchase on the subject matter. It would not enable you to defend the use of the law of cosines or its use in any particular context. On the other hand, if you understand how the

law of cosines figures in trigonometry, what it enables you to do, and how relying on it yields mathematically defensible results, your epistemic relation to it is different. You see that the requirement that trigonometry students master it is not an arbitrary demand, but rather is integral to understanding the subject. This is not to deny that the heteronomous demand can provide a stronger incentive. The desire to get out of 10th grade carries considerable weight. But the student's relation to the subject matter is different if she understands why the various expectations are placed on her and how realizing those expectations fosters her epistemic agency.

Not surprisingly, educating for epistemic autonomy involves treating students with respect. Rather than thinking of them as unformed clay to be molded into what their elders and betters consider educated persons, it treats them as agents, or at least proto-agents, capable of giving and appreciating reasons for what they think, and capable of recognizing that viewpoints that cannot be supported by reasons are untenable. Initially, of course, the reasons will be fairly simple. But with the growth of understanding, both a student's reasons and her capacity to give and appreciate reasons are refined and deepened.

Virtue redux

It might seem that I have simply replaced virtue responsibilism with deontology, that I have supplanted the virtues with the epistemic imperative. This is not quite right. On the position I have sketched, the epistemic imperative underwrites the epistemic virtues in that it is what explains and justifies calling certain dispositions virtuous. In that sense, it is basic. But the virtues underwrite the epistemic imperative as well. Strictly, the imperative requires only behaviors that *accord* with the virtues, not behaviors that *display* the virtues. An agent who lacked the disposition of fair-mindedness could on occasion act fair-mindedly anyway. But the stability of the realm of ends requires that its members regularly comply with the epistemic imperative and that they can count on their compatriots doing so as well. If compliance were sporadic, there could be no realm of epistemic ends. If an individual's compliance were sporadic, he would not be a responsible agent. Even if on a given occasion an intellectually fickle agent advocated a contention he reflectively endorsed, given his character, he could just as well have behaved heteronomously. Others would have no reason to trust him; indeed he would have no reason to trust himself. The dispositions that qualify as epistemic virtues are thus needed to steady the character of the individual agent and stabilize the realm of ends. Compliance need not be exceptionless, but failures to comply must be relatively rare. Members of the realm of ends reflectively endorse the epistemic virtues because they recognize that the exercise of those virtues leads to the formation and sustaining of beliefs they can reflectively endorse. The relation between the epistemic virtues and the epistemic imperative is reciprocal.

Acknowledgements

I am grateful to A. R. Mudd, Kenneth Walden, and the participants at the 2012 Spencer Foundation Colloquium on Virtue and Education for valuable comments on and criticisms of an earlier draft of this article.

Notes

1. I do not assume that only skeptical scenarios are epistemically unfortunate. There are what Jonathan Vogel has called semi-skeptical situations, where we know considerably less than we think we do. His discussion suggests that the actual world might be one where semi-skepticism obtains (Vogel, 1990). There are, moreover, cases where, through no fault of the agent's, the evidence is skewed or misleading. So the best thing to think on the basis of the evidence is relatively far from the truth.
2. Agents are responsible for actions they do when drunk or under the influence of recreational drugs. As Aristotle argues, even if they are out of control when they perform the particular actions in question, because were in control when imbibed to the point where they lost control, they are responsible for their drunken or drugged behavior. The bearing of this point on my argument is perhaps this: If someone irresponsibly neglects to exercise or acquire an epistemic virtue, he can be held responsible for the epistemic defects that result.
3. Thanks to A. R. Mudd for encouraging me to clarify this point.
4. Such heteronomy is characteristic of what Sosa (2007) calls animal knowledge. It is plausible that animals cannot reflectively endorse or criticize their attitudes, and are subjectively vulnerable in just the way I suggest. On the other hand, they presumably do not go in for, and do not need to go in for, self-reflection. So the vulnerability is not a subjective problem for them.
5. I think that epistemic agents can hone their perceptual, representational, and imaginative capacities. So they are not entirely passive even with respect to inputs. But I will not argue for that here.
6. This is a variation on the Molyneux problem (see Locke 1959 [1694], book 2, chapter 9, section 8).
7. I add 'displayable' to allow reasons that can be conveyed in non-propositional formats such as diagrams, pictures or maps.
8. See Adler (2002: 74–101), for an argument that coherence constraints are far more demanding than we might suppose.

References

- Adler J (2002) *Belief's Own Ethics*. Cambridge, MA: MIT Press.
- Axtell G (1997) Recent work in virtue epistemology. *American Philosophical Quarterly* 34: 410–430.
- Baehr J (2011) *The Inquiring Mind*. Oxford: Oxford University Press.
- Elgin C (2011) Science, ethics, and education. *Theory and Research in Education* 9: 251–263.
- Locke J (1959 [1694]) *An Essay Concerning Human Understanding*. New York: Dover.
- Kant I (1981 [1785]) *Grounding for the Metaphysics of Morals* trans. JW Ellington. Indianapolis, IN: Hackett.
- Korsgaard C (1996) The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values. In: *Creating the Kingdom of Ends*, pp. 275–310. Cambridge: Cambridge University Press.
- Plato (1976 [390 BCE]) *Meno*, trans. GMA Grube. Indianapolis, IN: Hackett.
- Rorty A (1986) The two faces of courage. *Philosophy* 61: 317–330.
- Rowling JK (1999) *Harry Potter and the Prisoner of Azkaban*. New York: Scholastic Press.
- Sosa E (2007) *A Virtue Epistemology*. Oxford: Clarendon Press.
- Vogel J (1990) Are there counterexamples to the closure principle? In: MD Roth and G Ross (eds) *Doubling: Contemporary Perspectives on Skepticism*, Norwell: Kluwer.
- Williams B (1981) Internal and external reasons. In: *Moral Luck*, pp. 101–113. Cambridge: Cambridge University Press.

Author biography

Catherine Z Elgin is Professor of the Philosophy of Education at Harvard Graduate School of Education. She is the author of *Considered Judgment* (1996), *Between the Absolute and the Arbitrary* (1997), *With Reference to Reference* (1983), and co-author with Nelson Goodman of *Reconceptions in Philosophy and Other Arts and Sciences* (1988). She is editor of *The Philosophy of Nelson Goodman*, a four-volume collection, and co-editor with Jonathan E. Adler of *Philosophical Inquiry*. She works on issues at the intersection of epistemology, philosophy of science, and philosophy of art.