

VIRTUE EPISTEMOLOGY AND THE ACQUISITION OF KNOWLEDGE

Duncan Pritchard

The recent literature on the theory of knowledge has taken a distinctive turn by focusing on the role of the cognitive and intellectual virtues in the acquisition of knowledge. The main contours and motivations for such virtue-theoretic accounts of knowledge are here sketched and it is argued that virtue epistemology in its most plausible form can be regarded as a refined form of reliabilism, and thus a variety of epistemic externalism. Moreover, it is claimed that there is strong empirical support in favour of the virtue epistemic position so understood, and an empirical study regarding the cognitive processes employed by medical experts in their diagnosis and treatment of epilepsy is cited in this regard. In general, it is argued that one can best account for 'expert' knowledge in terms of a virtue-theoretic epistemology that retains key reliabilist features. It is thus shown that understanding knowledge along virtue-theoretic lines has important implications for our understanding of how knowledge is acquired, and thus for the philosophy of education.

1. Process Reliabilism, Virtue Epistemology, and the Externalism/Internalism Divide

One of the defining debates in epistemology in the last 40 years has been the debate between epistemic internalists and externalists. This conflict represents far more than a clash of theoretical positions, since at the heart of this debate are two very different conceptions of what the epistemological project consists of. Moreover, since this debate has important ramifications for how one understands how knowledge is acquired, the implications of this conflict are clearly directly relevant to the philosophy of education.

For internalists the primary epistemic goal is the validation of one's judgements about one's epistemic position using only reflectively accessible resources (hence the internalist preoccupation with the problem of radical scepticism). Here epistemology is *ego-centric*, in that it starts with the reflective subject and moves outwards from this point, and also *intellectualist*, in that the possession of knowledge makes significant intellectual demands on the agent (only those subjects with the relevant reflective cognitive capacities can count as knowers). One consequence of these ego-centric and intellectualist aspects of internalism is that they significantly limit the import of the cognitive sciences to epistemology. Consider, for example, the empirical research conducted by psychologists on the cognitive processes underlying the formation of beliefs in certain subject matters. If the determination of one's knowledge is essentially dependent upon factors that are reflectively accessible to one, then it follows that these empirical facts about our cognitive

processes—facts which, presumably, stand outside of one's reflective ken—will not be relevant to the issue of whether or not one has knowledge (though they might of course still be epistemologically important in other respects).

In contrast, externalists are less concerned with ego-centric validation of this sort, and are similarly suspicious of the related intellectualist idea that the possession of knowledge should be essentially linked to the reflective capacities of the subject in the manner demanded by internalism. Without the ego-centricism the externalist is able to turn more to the cognitive sciences for assistance in the epistemological enterprise. In particular, empirical facts about how, for example, one's belief was formed—regardless of whether these facts are available to the subject—can be relevant to the possession of knowledge. Relatedly, moving away from the intellectualism of internalism broadens out the range of agents who can count as knowers by weakening one of the prerequisites for knowing.¹

Nowhere is this clash of philosophical viewpoints more transparent than in the debates over reliabilism, especially the bare form of process reliabilism famously advocated by Goldman in the mid-1960s.² Simplifying somewhat, the process reliabilist claim is that to know is to form a true belief via a process which is in fact reliable, where this means that the process tends to produce a sufficiently high number of true beliefs relative to false ones. And note what is key to this formulation, since the claim is precisely *not* that knowledge results when the agent forms a true belief via a process that she has good reflectively accessible grounds for believing is reliable, but only when the process is, in fact, reliable. This leaves open the possibility that an agent might have knowledge because she forms her true belief in a reliable fashion even though she lacks good reflectively accessible grounds for thinking that she is exhibiting such reliability.

This approach clearly indicates a shift away from an internalist ego-centric approach to epistemology, since if one can have knowledge even while lacking good reflectively accessible grounds in favour of that knowledge, then it cannot be regarded as part of the epistemic enterprise that it consists of an account of how the agent legitimates her own knowledge, since knowledge can be possessed in the absence of a legitimation of this sort. Similarly, the intellectualism inherent to internalism is here absent since it is not a pre-requisite for counting as a knower on this view that the agent is in possession of substantial intellectual capacities (all that are required are those intellectual capacities necessary for the formation of the target belief).

This debate between externalism (here in the guise of process reliabilism) and internalism clearly has important ramifications for education theory, since these two camps are offering very different conceptions of what is required of an agent in the acquisition of knowledge. In order to see the externalist aspects of process reliabilism in more detail, consider the much-discussed case of the chicken-sexer. This is an agent who, by being raised around chickens, has over time acquired a highly reliable sense of smell which enables her to reliably distinguish between male and female chicks. Crucially, however, the story goes that chicken-sexers tend to have false beliefs about how they are doing what they are doing, supposing themselves to be touching or seeing something distinctive, even though there is in fact nothing distinctive for them to see or touch. Moreover, we can imagine an extreme case in this regard of a 'naive' chicken-sexer who not only has mostly false beliefs about how she is doing what she is doing but also lacks good reflectively accessible grounds for thinking that she is reliable in this respect (she has never checked her success ratio, say, but merely assumed that she is forming her beliefs in this regard reliably). To make the case as clear as possible, suppose further that there are no defeaters available

to the agent which should prompt her to doubt her abilities in this regard, such as the testimony of someone authoritative who says that there's no such thing as 'chicken-sexing'.³ Does such an agent have knowledge of what she believes?⁴

This is the juncture at which internalist and externalist intuitions tend to come apart. While the externalist will be inclined towards a positive answer to this question, the internalist will tend to demur. It is interesting to note the kinds of considerations that will be offered by each side in favour of their position. Externalists, for example, will tend to claim that the true beliefs being formed by the agent reflect a genuine cognitive achievement, in that these true beliefs are the result of her using one of her naturally formed reliable abilities. If this is right, then it seems perverse to deny such true beliefs the status of knowledge. Internalists, in contrast, will emphasise the fact that the agent is unable to offer sufficient reflectively accessible grounds in favour of her belief, and thus that she cannot epistemically legitimate her belief, and will use this consideration to motivate support for the claim that the agent lacks knowledge in this case. The contrast here between the ego-centric approach of the internalist and the very different non-ego-centric approach of the externalist is clear. On the externalist view, one does not have to vouch for one's epistemic position using only the epistemic resources that one has most immediately (i.e. reflectively) to hand. Instead, the security of one's epistemic position could be vouchsafed by factors that are beyond one's ken, such as the fact that one's belief resulted from reliable testimony from an expert, or from a reliable cognitive faculty.

Moreover, the fact that the internalist account imposes a far more onerous intellectual burden on an agent before she can count as a knower is also manifest. According to the externalist, agents who are able to make the necessary discriminations in order to count as having beliefs—and this will mean having the intellectual sophistication to at least possess the conceptual repertoire to think the relevant thought associated with the content of the belief—will be in the market for knowledge, regardless of how intellectually unsophisticated they may be in other respects. Whether this extends knowledge out beyond the human to animals (like high-order mammals) is contentious, but it certainly extends it out to those unreflective agents who would otherwise be denied knowledge, such as chicken-sexers.⁵

One further feature of process reliabilism which is salient for our purposes is how this treatment of knowledge has clear empirical implications for the cognitive sciences. The question of whether an agent's belief counts as knowledge is far more straightforward on this model since one simply needs to examine empirically the reliability of the belief-forming mechanism in play.⁶ Moreover, giving agents concrete advice on how to acquire knowledge is also relatively straightforward. Once reliable forms of belief-formation are identified—and this will mean also identifying the circumstances under which the process in question is reliable, along with those circumstances where the reliability is undermined—then epistemic advice to an agent on how to gain knowledge reduces to concrete directions on which processes the agent should employ (or try to employ) in forming one's beliefs.⁷

Clearly, however, a bare process reliabilism cannot be the full story as to what is needed in a theory of knowledge. After all, not just *any* reliable process is knowledge-conducive. Imagine, for example, someone who is using a thermometer to find out the temperature of a room. Moreover, suppose further that this thermometer is broken so that its reading is fluctuating randomly throughout the day within a certain range. On the face of it, using such a thermometer to find out the temperature of the room cannot

be a way of gaining knowledge about what the temperature is. The problem is, however, that we can adapt the example to preserve the reliability of the method of belief formation, even while leaving the thermometer malfunctioning. All we need is someone hidden in the room who is adjusting the temperature so that it corresponds with what the thermometer is reading whenever the agent looks at the thermometer and forms the relevant belief. In these circumstances the agent will be forming a true belief in a reliable fashion, in that, across a wide range of circumstances, where she forms this belief it will be true.

One of the intuitions guiding the thought that the agent cannot be counted as a knower in this case, even if her true belief is reliably gained, is that the truth of her belief does not reflect a genuine cognitive achievement of hers. In short, it is not that her beliefs are tracking the truth, but rather that the truth is tracking her beliefs. The intuition is that whatever knowledge is, it needs to reflect an ability on the part of the agent to form beliefs in ways that are sensitive to the truth, and mere reliability alone won't suffice for this.

It is cases like this which have prompted externalists—such as Sosa (1991) and, later, Greco (e.g. 1999)—to refine the bare process reliabilist thesis along virtue-theoretic lines and propose a version of reliabilism known as *agent reliabilism*.⁸ In essence, virtue epistemology, or early forms of it at any rate, are simply process reliabilist views which restrict the class of reliable processes that can count as knowledge-conducive. In particular, their claim, in essence, is that knowledge results when an agent gains a true belief from a reliable process which is a stable part of the cognitive character of that agent—i.e. the belief-forming traits, or epistemic *virtues*, broadly speaking, of the agent.

In order to see the point of such a modification of the basic process reliabilist thesis, one needs to reflect on the difference this change makes when it comes to dealing with cases such as the 'thermometer' case just described. In this example the agent is not counted as gaining knowledge by the lights of an agent reliabilist thesis, even though the process employed is, *ex hypothesi*, reliable. This is because the reliability in question has nothing whatsoever to do with the agent's cognitive character but rather simply reflects a fortuitous interference on the part of a further feature of the world. That is, the facts in this case are, critically, tracking the agent's belief rather than *vice versa*, and the possibility that reliability of this sort could lead to knowledge is ruled out by agent reliabilism. The point generalises: to gain knowledge one must do more than simply form one's true beliefs in a way that is reliable; one must instead reliably form one's true beliefs via the stable cognitive processes that make up one's cognitive character.

In effect, one can regard agent reliabilism as re-introducing the epistemic importance of the agent while at the same time avoiding the pitfalls associated with the ego-centric approach to knowledge that we saw inherent in internalist epistemology. Knowledge can still be gained in ways that are, in the relevant respect, beyond the ken of the agent, but this does not mean that knowledge acquisition is entirely divorced from salient features of the agent's cognitive character. This is the key point about the move to a virtue-theoretic account of knowledge. What is attractive about knowledge is that it is a cognitive achievement on the part of the agent, and this means that the agent is reliably gaining true beliefs in ways that reflect positively on her cognitive character. That is, mere reliability is not enough for a cognitive achievement—instead, the reliable true belief must also reflect a cognitive ability on the part of the agent, where this means that the agent displays a cognitive *virtue*, broadly construed.

Crucially, however, agent reliabilism retains the key externalist element of process reliabilism since it continues to maintain that the features of the agent's cognitive character

which ensure that she gains true beliefs in a reliable fashion need not be reflectively available to the subject, and thus that even entirely unreflective agents can qualify as knowers. In short, for agent reliabilists, as for process reliabilists, knowledge can sometimes be acquired in a completely unreflective way. To return to the example given earlier, even naïve chicken sexers can have knowledge of what they unreflectively believe just so long as they are forming true beliefs in a reliable fashion as a result of the cognitive traits that make up their cognitive character (which, *ex hypothesi*, they are).

2. General Empirical Support for Agent Reliabilism

There is a great deal of empirical support for an agent reliabilist conception of how we gain knowledge. In particular, there have been a number of studies on cognition which point to how knowledge is gained by the use of belief-forming traits which are in the relevant sense sub-personal, in that they are employed by the agent in an entirely tacit fashion. This supports the agent reliabilist view in that it indicates that knowledge can be gained by agents just so long as they employ cognitive traits that are stable features of their cognitive character, even if these methods of belief-formation are not accompanied by the agent holding reflectively accessible grounds in favour of the target beliefs. As Greenwald (1992, 801) has commented, 'most of the "real work", both in the acquisition of cognitive procedures and skills and in the execution of cognitive operations, such as encoding and interpretation of stimuli, is being done at the level to which our consciousness has no access'. If this is right, and if we are to treat such cognitive activity as genuinely knowledge-conducive (we will return to this last point in a moment), then some form of agent reliabilism will be essential to understanding this knowledge, since cognitive activity of this sort could never count as knowledge-conducive by the lights of internalist accounts.

Q1

There are a number of factors that speak in favour of understanding cognition along agent reliabilist lines which bear on the apparent 'tacit' or unconscious nature of cognition. Three such factors stand out. The first is that agents have been shown to be able to respond to stimuli—and in responding, *process* that stimuli—in cases where the stimuli in question is far too complex to be understood as featuring at the level of the agent's conscious thought (see Broadbent 1993; Cleeremans and McClelland 1991; Lewicki, Czyzewska, and Hoffman 1987). Relatedly, it has been shown that agents can process kinds of stimuli that do not lend themselves to explicit conscious representation—such as very subtle variations in human appearance and behaviour (see Crain and Crain 1987; Lewicki 1986); the temporal patterns underlying trains of events (Michon 1990); and base rates (Christensen-Szalanski and Bushyhead 1981; Tversky and Kahneman 1974).

Q2

The second factor is the existence of 'implicit learning', where an agent acquires cognitive skills without engaging reflectively with the cognitive practice in question at all. An obvious case in this respect is the kind of 'hard-wired' learning one finds discussed in developmental psychology—the example of how a (normal) child learns to read is an obvious case in this regard. Here the agent lacks the capacities which would enable her to reflect on her learning in the first place, but this does not impede the learning process (see LaBerge and Samuels 1974). More generally, it has been shown that agents can acquire a capacity to recognise patterns without being conscious of what they have learned, and so such 'implicit' learning is not solely the province of developmental psychology—even agents capable of fully reflective thought can acquire cognitive skills in this implicit fashion (see Lewicki, Czyzewska, and Hoffman 1987; Reber 1989).⁹

Finally, the third factor concerns the importance of 'default reasoning' when it comes to forming judgements. Usually, default reasoning involves the largely unreflective use of 'rules of thumb' to form judgements about a certain subject matter, though in more extreme instances it may not involve any reasoning as such at all, as in cases where the agent unreflectively forms an immediate judgement in response to a specific stimulus. As epistemologists and those working in artificial intelligence have noted, the increases in reliability gained by greater computational complexity in the formation of a judgement may well not suffice to off-set the computational costs involved, such as the greater processing time (see, for example, Bach 1985). In general, a number of cognitive scientists have noted that there is good empirical evidence for thinking that our cognitive capacities have evolved in such a way as to enable us to make use of simple heuristics which 'short-circuit' complicated computational tasks without thereby losing an undue degree of reliability in our judgements (e.g. Gigerenzer and Todd 1999). The acquisition of knowledge via the acquisition of reliable true belief will thus in a wide range of cases be achieved in a largely unreflective fashion.¹⁰

Still, one might be suspicious about what morals should be drawn from empirical research of this sort, since it has been argued by some proponents of virtue epistemology who have opted for an internalist rendering of the thesis—such as Zagzebski (1996, section 2.2), who describes her view as 'neo-Aristotelian'—that knowledge possession demands reflective engagement, and thus that those who form their beliefs in an insufficiently (or inappropriately) reflective fashion, such as young children or naïve chicken-sexers, for example, cannot count as knowers.¹¹ One motivation that Zagzebski offered in favour of this claim is that this understanding of what a virtue epistemology involves is more in keeping with the modern sense of 'virtue'. To the contemporary ear, this term refers (if it refers at all to intellectual or cognitive virtues) to those intellectual or cognitive virtues which are explicitly reflective, like the virtue of being open-minded, or conscientious. In each case, one would expect such a virtue to manifest itself in the agent reflecting on her practices in ways that produce supporting grounds for the practice that are reflectively available to the subject. The conscientious agent, for example, will tend to be an agent who is careful in how she is forming her beliefs and thus seeks additional supporting grounds in favour of those beliefs, grounds which she can cite if called upon to do so.

Q3

Of course, agent reliabilists can allow that there are cognitive virtues of this sort; their claim is only that there are also cognitive virtues that do not involve reflection in this way. Indeed, where the agent reliabilist is sensitive to the recent empirical data from the cognitive sciences, she will typically maintain that it is essential to the view that it does not impose this demand, since there are a wide range of cases in which agents employ stable cognitive traits to reliably form their true beliefs, and yet lack good reflectively accessible grounds to back up the beliefs so formed. Intuitively, however, such cases do count as instances of knowledge, contrary to the claims of the more demanding group of virtue epistemologists who cast their view along internalist lines.¹²

In order to deal with this issue head-on, we will focus on a particular empirical study in detail.

3. Specific Empirical Support for Agent Reliabilism—Expert Medical Diagnosis

The study that we will focus on concerns the methods of belief-formation employed by expert medical practitioners when it comes to offering medical diagnoses. One

advantage of making use of an example of this sort is that it should be uncontroversial that the experts in this regard are at least capable of the kind of reflective engagement demanded by some as a necessary condition for knowledge possession. Moreover, as experts who are—and we can empirically verify this claim of course—at the peak of their field in this regard, there cannot be any serious concern that the diagnostic strategies being employed are sub-standard (though this is not to say, of course, that they cannot be improved upon). By definition, the belief-forming processes employed by these doctors represent ‘best practice’ in the profession.

One feature of expert medical diagnosis that has puzzled cognitive scientists has been how the diagnostic strategies employed by medical experts rarely advert to the explicit rules that medical practitioners are taught in their training, but instead reflect rules of thumb which are often not explicitly formulated even by the expert herself. A good example of this comes from the area of pharmacology and concerns medical diagnosis and treatment in the case of epilepsy—especially when the anti-epileptic drug phenytoin is being prescribed. This is a complex affair since there are a number of relevant factors that need to be determined before an anti-epilepsy therapy that is optimum for the patient concerned can be chosen. In particular, the doctor needs to (i) classify the epileptic seizure type; (ii) determine whether or not the seizure type in question requires anti-epileptic drugs; and (iii) determine what dosage of phenytoin, where applicable, should be administered to the patient. This last aspect of therapy choice is particularly tricky, since the correct dosage needed will depend upon, amongst other things, (i) the type of seizure; (ii) the patient’s age; (iii) the patient’s weight; (iv) the medical history of the patient; (v) the patient’s behaviour (are there any signs of overdosing?); and (vi) the possibility of interaction between this drug and other drugs that the patient may be currently prescribed. Moreover, there is a further constraint on correct therapy where phenytoin is administered. On the one hand, the doctor needs the initial dosage to be relatively strong in order to ensure that it has an effect so that the patient doesn’t suffer from repeated seizures. On the other hand, however, the doctor also needs to be wary of administering too strong a dosage and causing the patient to overdose.

A number of studies have been conducted that try to determine the kinds of methodologies employed by expert practitioners in this regard, as opposed to relative novices (such as final-year medical students). Part of the goal of this research is to determine what constitutes best practice in this regard (and verify that the ‘best’ practice really is superior to the other diagnostic methods available) in order to aid the teaching of medical practitioners. We will focus here on one such study, due to Boreham, Foster, and Mawer (1992).¹³ This study looked at the diagnostic practices of three groups: expert practitioners, final-year medical students (all of whom had taken courses in pharmacology which dealt with the prescription of anti-epileptic drugs like phenytoin) and research students in pharmacology.

As expected, the expert practitioners consistently made optimal decisions regarding the dosage of phenytoin needed in each case, where the dose was strong enough to stop the seizures but not so strong as to cause an overdose in the patient. They were also suitably sensitive to the relevant factors in this regard, such as the patient’s medical history. Of the final-year medical students in the second group, around half of them opted for an optimal dosage, with the other half roughly split between those who offered a dosage which was too cautious and those who offered a dosage which was large enough to be hazardous.¹⁴ Finally, the students in the third group—the research pharmacologists—all opted for non-optimal dosages, and frequently the dosages they proposed were large enough to be hazardous.

Interestingly, the members of the three groups employed very different diagnostic strategies. The worst performing group—the research pharmacology students—tended to proceed by first looking up the range of effective plasma concentrations in the *British National Formulary* (a reference work). Phenytoin works by increasing the plasma concentration in the patient's blood plasma and the diagnostic goal is to gain a steady plasma concentration at which the seizures stop but at which the agent is not suffering the hazardous effects of a dosage which is too strong. Since this reference work states that the effective plasma concentration is between 10 and 20 mg per litre of blood, depending on the individual patient, students in this grouping tended to take the average effective plasma concentration of 15 mg per litre as their target and to estimate the dosage needed to take the patient's plasma concentration to this level. Since this average will be too high for some patients, and since the students in this group were insufficiently sensitive to other relevant variables (such as the medical history of the patient), this was the reason why the members of this group tended to opt for dosage levels that were hazardous.

In contrast, students in the second group—which was composed of final-year medical students—tended to opt for a more cautious strategy which consisted of simply adding the smallest possible increments to the dosage until either the seizures stopped or they reached the recommended upper-limit of doses as set out in the *British National Formulary*. Although this strategy had the advantage of not leading to overdoses, it did mean that often patients were under-prescribed the drug they needed, and this left them vulnerable to further seizures. Again, part of the problem here was that the members of this group lacked the expertise to be sufficiently sensitive to other facts which were salient to forming the right diagnosis, such as relevant facts about the patient.

Finally, the experts consistently offered the optimal dosage, one which was neither too little, thereby leaving the patient open to further seizures, nor so large as to expose the patient to the hazardous effects associated with overdosing. What was interesting about the expert strategy was that it did not conform to the recommended rules for treatment that are taught to medical students (the final-year medical students were employing a practice that roughly corresponded to these rules). Instead, the experts were far more sensitive to the concrete facts of the situation—such as regarding the patient's demeanour and case history—and employed rough-and-ready rules of thumb which were different in detail to the rules set out in the teaching of medical students in this regard. One of these rules was the following, as stated by one of the experts:

As a rule of thumb, if plasma concentration is less than 7 mg/l, an increase of 100 mg a day is safe. Between 7 and 10 mg/l, an increase of 50 mg is safe. Over 10 mg/l, one would want to edge up in 25 mg increments. (Boreham, Foster, and Mawer 1992, 26)

Although all the experts appeared to be using a rule of this sort, only one of them was able to state the rule explicitly in this way. Moreover, it is significant that this rule does not correspond to anything that is formally taught in medical colleges, even though, as one would expect, the rule is generally in keeping with what is known about the relationship between phenytoin dosages and blood plasma concentrations. As Boreham, Foster, and Mawer put the point:

The decreasing size of the dose increments specified in these rules reflects the curvilinear relationship between the daily dose and steady-state plasma concentration defined by the modified Michaelis-Menten equation. It is possible that knowledge of the general

shape of this function guided the experts in compiling their rules out of clinical experience. However, the specific rule content appears to have been drawn mainly from clinical experience, for there is nothing in pharmacokinetic theory which would identify 7 mg/l and 10 mg/l as the cut-off points for making the step down to the next size of safe dose increment. (1992, 26)

It seems then that best practice in this specific case of medical diagnosis and treatment consists of agents following tacit belief-forming methods which are (i) acutely sensitive to concrete features of the situation and (ii) make use of rules of thumb which even the expert herself is often unable to articulate fully (and which are unlikely to have ever been explicitly articulated).

This result is important, since it appears to back up the general tenor of agent reliabilism as a virtue-theoretic account of knowledge which, in line with other externalist treatments of knowledge, does not demand of the agent that she is able to offer sufficient reflectively accessible grounds in favour of her beliefs before she can count as a knower. For while it is no doubt true that reflection in these cases may well, in part at least, improve best practice (this is one of the points of the research, to make the processes employed in best practice explicit so that they can be taught to medical students¹⁵), it does not seem to be necessary to insist on such reflection in order to count the experts in question as knowing what the correct diagnosis is. In particular, what seems to be important is that these experts are employing reliable cognitive mechanisms in a stable fashion which is duly sensitive to relevant features of the environment. If they are able explicitly to formulate the precise ways in which their belief-forming process is informed by these factors—i.e. by being able to specify the rules that they are following, and list the concrete features of the situation that they are responding to, along with the relative weight that they are according to them—then this is all well and good, though not, it seems, essential for knowledge possession in this regard.

Indeed, in one very important sense of the term 'virtue', these experts are clearly manifesting intellectual virtue since they are exhibiting an *excellence* of cognitive character, albeit one that is not always accompanied by reflection on the character traits in question. Furthermore, it is important to note that this virtue-theoretic form of reliabilism has the advantage over more simple process reliabilist theories in that by making the connection to the agent's cognitive character explicit, it is clear that the cognitive traits under consideration are of epistemic credit to the agent. That is, in common with all knowledge, the agent is forming a true belief in a non-lucky fashion in a way that is of credit to that agent. This is in marked contrast to the kinds of 'lucky', albeit reliably formed, true beliefs that often feature in counterexamples to simple reliabilism, such as in the 'thermometer' case described earlier.¹⁶

4. The Epistemology of Expertise

This point about the epistemology of expertise in the case of medical diagnoses is borne out in a number of other discussions on expert knowledge. For a common theme in studies of expert knowledge is how the knowledge employed is tacit and thus not capable of codification even by the agent herself. In the cases of medical expertise that we just considered, there was knowledge that could be made explicit and then be used to inform the training of others—namely, the rule of thumb the experts used in their

diagnosis. Note, however, that there was a great deal else in play when the medical experts formed their diagnoses than simply this rule. Moreover, these other aspects of the diagnostic activity, such as being sensitive to facts about the patient's demeanour or her medical history which might seem insignificant (or just not noticed at all) by the novice, are not obviously codifiable in terms of a set of rules. Crucially, however, this is not a feature solely of expert medical diagnosis, but instead reflects expert practice in a wide range of cases.

This point has, for example, been made by Luntley (2004). He considers the case of a teacher who 'is concerned with managing some feature of classroom environment, for example, keeping noise levels within a range that makes learning both fun and productive' (Luntley 2004, 15). As Luntley points out, there is 'no rule, definable in terms of individual events and behaviour, that defines what an appropriate overall noise level is' (2004, 15). Indeed, given the complexity of the task in hand, and the myriad contextual factors which are relevant to making a decision as to what constitutes acceptable overall noise level (facts concerning, for example, the tasks in hand and the psychological make-up of the students involved), it would be absurd to suggest that what constitutes expert knowledge in this case is some sort of implicit theoretical appreciation and processing of the salient facts. As Luntley neatly puts the point:

Even in this very simple example, it would be heroic to insist that teacher expertise consisted in the internalisation of a theory of how classes make noise; what sorts of noise are appropriate to learning and what are not; which sorts of children can be left to occasionally 'boil over' in noise production and which have to be much more closely monitored; which sorts of rooms will tolerate higher noise levels and which sorts will reverberate, etc. (2004, 16)

Expertise here clearly does not consist—*nor could it consist*, given the general cognitive limitations of the human animal—in the ability to back up one's reliable judgements about the target proposition (in this case what the appropriate noise level should be) by sufficient reflectively accessible grounds. And yet clearly the teacher can be an expert at making judgements of this sort (it is certainly not by chance that she is reliable in this respect), and she is surely not only exhibiting reliability in making these judgements, but also displaying a cognitive character which is epistemically virtuous in just the relevant respects needed to reliably make judgements of this sort. Expertise, then, whether it be in the form of rather complicated medical diagnosis or the apparently simpler case of making judgements about appropriate noise levels in class, is best thought of in terms of the exercise of epistemic virtues on the agent reliabilist model than in the internalist terms suggested by Zagzebski and others.¹⁷

5. Concluding Remarks

The move to the specific virtue-theoretic theory of agent reliabilism in contemporary epistemology is thus extremely significant to how we understand the process of knowledge acquisition since it represents an important and decisive shift away from the ego-centric and intellectualist epistemic project found in internalist epistemological theorising. Crucially, however, it makes this move without undermining the importance of the role of the agent and the agent's cognitive character in the production of *bona fide* knowledge, and in this sense it can be seen as a tempering of the early externalist process reliabilist

thesis. What is most salient for our purposes, however, is that the conception of knowledge embodied in agent reliabilism receives empirical support from recent work in the cognitive sciences by allowing knowledge to be possessed even when the agent is unable to offer adequate reflectively accessible grounds in favour of her belief. That such unreflective knowledge is often gained even in cases of 'best' epistemic practice—as in the case of the medical experts that we considered—indicates that talk of cognitive and intellectual virtues should not lead us back into the intellectualised model of knowledge that is distinctive of the internalist epistemological project, and which is currently being advocated by proponents of a very different style of virtue-theoretic epistemic thesis that is cast along internalist lines. We acquire knowledge by acquiring the reliable cognitive traits that make up our cognitive character. While this is consistent with knowledge possession sometimes demanding that the agent has adequate reflectively accessible supporting grounds available to her, crucially it is also consistent with no such grounds being available.¹⁸

The ramifications for the philosophy of education should be clear. While the acquisition of knowledge through learning—whether in infants or mature adults; the uninitiated or the expert—will certainly demand the possession of relevant cognitive traits on the part of the agent, it need not thereby demand that relevant *reflective* cognitive processes are in operation or even that such reflective processes are available to that subject at that time. In short, knowledge acquisition is sometimes a *completely* unreflective matter.

ACKNOWLEDGEMENTS

I am grateful to Nick Boreham, Michael Luntley, and Finn Spicer for assistance in the writing of this paper. Thanks also to The Leverhulme Trust for the award for a Special Research Fellowship which has enabled me to conduct work in this area.

NOTES

1. For more on the externalism/internalism distinction in epistemology in general, see Kornblith (2001) and Bonjour and Sosa (2003).
2. The early Goldman view receives its most refined expression in Goldman (1986).
3. The issue of defeaters is one reason why the famous 'clairvoyant' counterexamples to reliabilism proposed by Bonjour (e.g. 1985, chap. 4) are problematic, as one typically is in the possession—or at least ought to be, if one is sufficiently careful in how one forms one's beliefs—of a standing defeater regarding the reliability of this belief-forming process. In particular, modern psychology is often characterised, especially in the public realm, in terms of an opposition to the supposed reliability of belief-forming processes of this sort. Once one notes the presence of the defeater, however, then the example starts to lose its dialectical force since it is typically part of the process reliabilist view, at least when refined, that an agent who does not suitably revise her belief in response to defeaters *can* as a result lack knowledge of what she believes even if the underlying process which gave rise to the belief was as a matter of fact reliable. See, for example, Goldman (1986, 62–63) and Greco (forthcoming). The key reliabilist point is that mere reliability alone can sometimes suffice for knowledge, not that it *always* suffices.
4. It isn't always clear when this case is being discussed just what the content of the target belief is supposed to be. Is it, for example, the belief that the chick which the agent is

holding is of a certain sex? This is possible, though I don't think the example should be saddled with quite so specific a belief content, and the usual way that the chicken-sexer's abilities are described tends to count against such a description in any case. Instead it seems that what is at issue is the epistemic status of the agent's belief that the chick which she is holding is of a different sex to some other specified chick. Having the ability to reliably tell male and female chicks apart is one thing; having the ability to reliably tell them apart *and* identify which set of chicks is male and which is female is more demanding.

5. For further discussion of the chicken-sexer example, see Foley (1987, 168–69), Lewis (1996), Zagzebski (1996, sections 2.1, 4.1), and Brandom (1998). See also the exchange between Sainsbury (1996) and Wright (1996). Note that it is entirely consistent with the externalist view that it grants that there is something epistemically desirable about being in a position to offer adequate reflectively accessible grounds in favour of one's beliefs. The point is simply that such a reflective capacity is not necessary for knowledge possession.
6. Some have argued that identifying which belief-forming mechanism is in play is itself a near-impossible task. This is the so-called 'generality' problem (see, for example, Conee and Feldman 1988). To my mind, however, this problem has been overstated. While it can obviously be sometimes very difficult to determine empirically which belief-forming mechanism an agent is employing, it is far from impossible (indeed, the cognitive science literature is full of studies which identify belief-forming mechanisms). For a subtle discussion of how reliabilists might meet the generality problem, see Goldman (1993).
7. Of course, it may be that the relevant process is one that the agent cannot master straight away, perhaps because it requires special training or abilities. In such cases, it may be that the optimum process in terms of reliability *for this agent at this time* is very different to the optimum process *simpliciter*. I return to this point below when I discuss 'expert' knowledge.
8. It is actually just Greco who describes his view as an agent reliabilist thesis, but the similarities between his view and Sosa's are strong enough to count them both as falling under this description for our purposes. See also the 'proper functionalist' theory presented by Plantinga (1993), which bears a number of key similarities to the basic agent reliabilist thesis. It should also be noted that Goldman now presents his version of reliabilism in a way that is broadly in line with agent reliabilism. See, for example, Goldman (1993).
9. For further discussion of these empirical studies, see Boreham (1994).
10. For more on this point, see Spicer (2004). For a useful survey of the main *a priori* and empirical considerations which count against internalist epistemologies, see Goldman (1999).
11. There is a third option in this respect, which is to allow that there might be two types of knowledge—one which is 'brute' and unreflective, and another which is reflective. This type of position is most usually associated with the work of Sosa (1995, 1997). For discussion, see Grimm (2001; cf. Sosa 2001) and Greco (2003, 298–301). For an overview of the two types of virtue epistemology—the early and broadly reliabilist and externalist view on the one hand, and the later and broadly *responsibilist* and internalist view on the other—see Axtell (1997) and Brady and Pritchard (2003).
12. See Sosa (1991, 271), for an explicit defence of the more inclusive conception of 'virtue'.
13. See also Boreham (1995) and Boreham, Foster, and Mawer (2000).

14. There is actually a further complication in this regard that we do not need to go into, which is that half of the final-year medical students had played a game called 'The Phenytoin Game' which is designed to aid correct diagnosis of the dosage of phenytoin needed. While playing this game didn't ensure that the students made the correct diagnosis, it did have an effect in that students who had played the game yet made sub-optimal diagnoses tended to opt for overly cautious, rather than hazardous, doses of the drug.
15. Note, however, that it is not immediately obvious that one should teach the more subtle rules followed by experts to medical students, since only the experts have the wealth of background knowledge available to them that can allow them to utilise these rules effectively. Indeed, the 'cautious' rules already followed by many final-year medical students in this regard may be better rules to follow to begin with. At the very least, though, it is useful to make this tacit knowledge explicit so that it can guide research and inform the education of medical students.
16. For more on the relationship between virtue epistemology and the issue of epistemic luck—i.e. the sense in which knowledge is non-lucky true belief—see the symposium between Axtell (2003), Greco (2003), and Pritchard (2003). For more on the topic of epistemic luck more generally, see Pritchard (2005). Recent studies, both in philosophy and cognitive psychology, have emphasised the importance of luck in understanding agent responsibility, whether cognitive or otherwise. For a critical survey of this literature, see Pritchard and Smith (2004).
17. Luntley discusses the status of expert knowledge in a number of works. See, for example, Luntley (2002). In general, his position in this regard relates to a novel theory of concept possession. I think that this project shares several key themes with the agent reliabilist conception of knowledge, though to argue for this in detail is a task for another occasion.
18. This is not to suggest, of course, that there are *no* problems of substance facing the agent reliabilist view, only that this thesis represents the most satisfactory theory of knowledge currently available. Pritchard (2003) discusses one key problem for virtue epistemology which concerns how strong the thesis should be formulated.

REFERENCES

- AXTELL, G. 1997. Recent work on virtue epistemology. *American Philosophical Quarterly* 34: 1–26.
- . 2003. *Felix culpa*: Luck in ethics and epistemology. *Metaphilosophy* 34: 331–52.
- BACH, K. 1985. A rationale for reliabilism. *The Monist* 68: 246–63.
- BONJOUR, L. 1985. *The structure of empirical knowledge*. Cambridge, Mass.: Harvard University Press.
- BONJOUR, L., and E. SOSA. 2003. *Epistemic justification: Internalism vs. externalism, foundations vs. virtues*. Oxford: Blackwell.
- BOREHAM, N. 1994. The dangerous practice of thinking. *Medical Education* 28: 172–79.
- . 1995. Error analysis and expert–novice differences in medical diagnosis. In *Expertise and technology*, edited by J.-M. Hoc, E. Hollnagel, and C. Cacciabue. Hove: Erlbaum.
- BOREHAM, N., R. W. FOSTER, and G. E. MAWER. 1992. Strategies and knowledge in the control of the symptoms of a chronic illness. *Le Travail Humain* 55: 15–32.
- . 2000. Medical students' errors in pharmacotherapeutics. *Medical Education* 34: 188–93.
- BRADY, M. S., and D. H. PRITCHARD. 2003. Moral and epistemic virtues. In *Moral and epistemic virtues*, edited by M. S. Brady and D. H. Pritchard. Oxford: Blackwell.
- BRANDOM, R. 1998. Insights and blindspots of reliabilism. *The Monist* 81: 371–92.

- BROADBENT, D. 1993. Planning and opportunism. *The Psychologist* 6: 54–60.
- CHRISTENSEN-SZALANSKI, J., and J. BUSHYHEAD. 1981. Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception of Performance* 7: 928–35.
- CLEEREMANS, A., and J. MCCLELLAND. 1991. Learning the structure of event sequences. *Journal of Experimental Psychology: General* 120: 235–53.
- CONEE, E., and R. FELDMAN. 1988. The generality problem for reliabilism. *Philosophical Studies* 89: 1–29.
- CRAIN, W., and E. CRAIN. 1987. The psychology of the scientist: Can humanistic theory contribute to our understanding of medical problem solving? *Psychological Reports* 61: 779–88.
- FOLEY, R. 1987. *A theory of epistemic rationality*. Cambridge, Mass.: Harvard University Press.
- GIGERENZER, G., and P. TODD. 1999. *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- GOLDMAN, A. 1986. *Epistemology and cognition*. Cambridge, Mass.: Harvard University Press.
- . 1993. Epistemic folkways and scientific epistemology. *Philosophical Issues* 3: 271–84.
- . 1999. Internalism exposed. *Journal of Philosophy* 96: 271–93.
- GRECO, J. 1999. Agent reliabilism. *Philosophical Perspectives* 13: 273–96.
- . 2002. Virtues in epistemology. In *The Oxford handbook of epistemology*, edited by P. K. Moser. Oxford: Oxford University Press. Q5
- . 2003. Virtue and luck, epistemic and otherwise. *Metaphilosophy* 34: 353–66.
- . Forthcoming. Truth in light of virtue. *Philosophical Studies*. Q6
- GREENWALD, A. 1992. Unconscious cognition reclaimed. *American Psychologist* 47: 93–104.
- GRIMM, S. 2001. Ernest Sosa, knowledge and understanding. *Philosophical Studies* 106: 171–91.
- KORNBLITH, H., ed. 2001. *Epistemology: Internalism and externalism*. Oxford: Blackwell.
- LABERGE, D., and S. SAMUELS. 1974. Toward a theory of automatic information processing in reading. *Cognitive Psychology* 6: 293–323.
- LEWICKI, P. 1986. *Nonconscious social information processing*. New York: Academic Press.
- LEWICKI, P., M. CZYZEWESKA, and H. HOFFMAN. 1987. Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition* 13: 523–30.
- LEWIS, D. 1996. Elusive knowledge. *Australasian Journal of Philosophy* 74: 549–67.
- LUNTLEY, M. 2002. Knowing how to manage: Expertise and embedded knowledge. *Reason in Practice: The Journal of Philosophy of Management* 2: 3–14.
- . 2004. Growing awareness. *Journal of Philosophy of Education* 38: 1–20.
- MICHON, J. 1990. Implicit and explicit representations of time. In *Cognitive models of psychological time*, edited by R. Block. Hillsdale, N.J.: Erlbaum.
- PLANTINGA, A. 1993. *Warrant and proper function*. Oxford: Oxford University Press.
- PRITCHARD, D. H. 2003. Virtue epistemology and epistemic luck. *Metaphilosophy* 34: 106–30.
- . 2005. *Epistemic luck*. Oxford: Oxford University Press.
- PRITCHARD, D. H., and M. SMITH. 2004. The psychology and philosophy of luck. *New Ideas in Psychology* 22: 1–28.
- REBER, A. 1989. Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General* 118: 219–35.
- SAINSBURY, R. M. 1996. Crispin Wright: *Truth and objectivity*. *Philosophy and Phenomenological Research* 56: 899–904.
- SOSA, E. 1991. *Knowledge in perspective: Selected essays in epistemology*. Cambridge: Cambridge University Press.

- . 1995. Perspectives in virtue epistemology: A response to Dancy and Bonjour. *Philosophical Studies* 78: 221–35.
- . 1997. How to resolve the Pyrrhonian problematic: A lesson from Descartes. *Philosophical Studies* 85: 229–49.
- . 2001. Human knowledge, animal and reflective. *Philosophical Studies* 106: 193–96.
- SPICER, F. 2004. Epistemic intuitions and epistemic contextualism. Unpublished manuscript.
- TVERSKY, A., and D. KAHNEMAN. 1974. Judgement under uncertainty: Heuristics and biases. *Science* 185: 1124–31.
- WRIGHT, C. 1996. Response to commentators. *Philosophy and Phenomenological Research* 56: 911–41.
- ZAGZEBSKI, L. 1996. *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge: Cambridge University Press.

Duncan Pritchard, Department of Philosophy, University of Stirling, Stirling FK9 4LA, UK.
E-mail: d.h.pritchard@stir.ac.uk

RPEX121934

Queries

Duncan Pritchard

- Q1 Greenwald (1992, 801): please check page number as it does not match the Reference list
- Q2 Broadbent 1989 changed to 1993 (as Ref. list) – OK?
- Q3 Zagzebski 1995 changed to 1996 (as Ref. list) – OK?
- Q4 Greco (2003, 298–301): please check page numbers as they do not match the Reference list (or should this be 2002?)
- Q5 Greco 2002: please cite in text or delete
- Q6 Greco forthcoming: any update?