

Forthcoming in INTELLECTUAL VIRTUE: PERSPECTIVES FROM ETHICS AND EPISTEMOLOGY, ed. M. DePaul and L. Zagzebski (OUP, 2002)

The Place of Truth in Epistemology¹

Ernest Sosa

...[Human] good turns out to be activity of soul in accordance with virtue, and if there are more than one virtue, in accordance with the best and most complete.

Aristotle, Nichomachean Ethics, Bk I, sec. 7.

... With those who identify happiness [faring happily or well] with virtue or some one virtue our account is in harmony; for to virtue belongs virtuous activity. But it makes, perhaps, no small difference whether we place the chief good in possession or in use, in state of mind or in activity. For the state of mind may exist without producing any good result, as in a man who is asleep or in some other way quite inactive, but the activity cannot; for one who has the activity will of necessity be acting, and acting well. And as in the Olympic Games it is not the most beautiful and the strongest that are crowned but those who compete (for it is some of these that are victorious), so those who act win, and rightly win, the noble and good things in life.

Ibid., Bk I, sec. 8.

... [Of] the intellect which is contemplative, not practical nor productive, the good and the bad state are truth and falsity respectively (for this is the work of everything intellectual).

Ibid., Bk VI, sec. 2.

I

In order to qualify as knowledge, a belief need only be both true and “apt.” What then is such aptness and what role might truth play in determining it? Is a belief (epistemically) apt insofar as it promotes some truth-involving goal? If so, which goal?

If knowledge *is* better than mere true belief, moreover, in what way is it better? How does our conception of epistemic aptness help explain why it is better to have an *apt* true belief than a mere true belief?

¹ My paper may amount to little more than a partial reading of these three passages by Aristotle (partial, perhaps, in more than one sense).

A belief does not count as apt simply because it promotes the goal of having true beliefs. A belief that a certain book is a good source of information may be ill-grounded and inapt though in fact true and, when acted upon, a source of much further true belief. A belief can promote a massive acquisition of true beliefs without thereby becoming apt.

We do well to replace that diachronic goal, therefore, perhaps with a synchronic goal of *now* acquiring true beliefs (and no false ones). But this threatens a *reductio*: that all and only one's present true beliefs will then be epistemically rational, by promoting one's goal of *now* acquiring true beliefs.

We might try replacing the simple synchronic goal with a subjunctive synchronic goal such as

G Being such that $(\forall x)(\text{One would now believe } x \text{ if and only if } x \text{ were true})$.

This avoids the *reductio*. Not every true belief is such that one *would* believe it only if true.

These truth goals nonetheless all share a problem: namely, how implausible it is to suppose that we either do or should have any such goal. We are, it is true, said to want the truth as intellectual beings. But what does this mean? It might mean that we want true beliefs, any true beliefs, since among the features that make a belief desirable is its plain truth. If so, is our time and energy always well used in acquiring true beliefs, *any* true beliefs? Is no true belief wholly ineffectual, even if we might then attain ends that we rightly value even more?

At the beach on a lazy summer afternoon, we might scoop up a handful of sand and carefully count the grains. This would give us an otherwise unremarked truth, something that on the view before us is at least a positive good, other things equal. This view I hardly understand. The number of grains would not interest most of us in the slightest. Absent any such antecedent interest, moreover, it is hard to see any sort of *value* in one's having that truth.²

Are we then properly motivated to acquire true beliefs simply under the aspect of their being true? More plausible seems the view that, for any arbitrary belief of ours, we would prefer that it be true rather than not true, other things equal. In other words,

- (a) so far as truth goes, we'd rather have it in any given belief that we actually hold.

However, this does not entail that

- (b) if all a belief has to be said for it is that it is true, then we prefer to have it than not to have it.

Nor does this follow even if we add that the belief is evaluatively neutral in every respect other than its truth.

To want the answer to a question, for its practical value or simply to satisfy our curiosity, is to want to know a truth. If I want to know whether p , for example, I want this: *to know that p , if p* , and *to know that not- p , if not- p* . And to want *to know that p , if p* , is to want *to know that it is true that p , if it is true that p* , and similarly for not- p . So our desire for truths is largely coordinate with our desire for answers to our various questions.

Just as we want the food we eat to be nutritious, so we want that the beliefs we hold be true, other things equal. Indeed, in pursuing the answer to a question we are automatically pursuing the truth on that question. But this does not mean we must value the truth *as* the truth, in the sense that, for any of the vast set of truths available, one must value one's having it at least in the respect that it is a truth. This no more follows for true propositions than does its correlate for nutritious food.

That distinction bears emphasis. I can want food that is nutritious, in this sense: that *if*, for whatever reason, because I find it savory, perhaps, I want to have—with my next meal, or just regularly and in general—bread, I would prefer that my bread be

² It might be replied that the value is indeed there though nearly indiscernibly slight. This I am not inclined to dispute, since what I have to say could be cast about as well in terms of vanishingly slight value, irrespective of whether its magnitude is epsilon or zero.

nutritious; which does not mean that I want, in itself and independently of its being food desired in other respects, that I have nutritious food simply for its nutritive value. In fact, of course, most of us do want regularly to eat nutritious food, as its own separable desideratum. Nevertheless, from (a) the premise that we want the food we eat to be nutritious rather than not, we cannot validly draw (b) the conclusion that we have a separable desire that we consume nutritious food, that we have an objective of next, or regularly, doing so, regardless of whatever *other* desires we may or may not have for sorts of food.

Similarly, we may want true beliefs, in this sense: that *if*, for whatever reason, we are interested in a certain question, we would prefer to believe correct rather than incorrect answers to that question; but this does not mean that we want, in itself and independently of our wanting these questions answered, for whatever independent reason, that we have true answers to them, simply for the truth this would give us.

What then of our belief formation? What do we hope for in that regard? Insofar as our belief formation is directed to answering questions we want answered, it is of course aimed at truth, trivially so, as we have seen. But which questions *should* we want answered, if any? Some questions we can hardly avoid: our very survival turns on them. Other questions we want answered for the sake of our comfort, and so on. Even once we put aside the most mundane questions, that still leaves a lot open. We shall be interested in a huge variety of questions, as family members, as citizens, and just as rational, naturally curious beings. Is there anything general to be said here? Can some general desire for the truth be recommended? It is hard to see what it could be. Remember, we have no desire for truths per se. When we have a desire for the truth, this is because that desire is implicated in our desire for an answer to a particular question or for answers to questions of some restricted sort. But our interest in the truth in such a case is just our interest in the question(s). If we can generalize beyond this to a recommendable desire for truth, accordingly, it must involve a generalization to a sort of questions that *should* draw our interest. But is there such a thing? Can we at least pick out a *sort* of truth that *should* interest us (apart from the sort “truths that should interest us” or variants of this)?

Your life goals may quite properly be different from mine once we move beyond the most abstract level of *living a good life* or *living a good life in the company of good fellow human beings in a good society*, or the like. Each of us may have such a goal, but great differences set in once we determine more specifically the shape of its realization in a life, given the constitution and context of that particular human being. Won't our intellectual goals be subject to this same kind of difficulty? Our interest in the truth is an interest in certain questions or in certain sorts of questions, and properly viewed as such. What questions interest a given thinker may properly differ, moreover, from those that interest others.

It might be replied that we do or at least should have these goals, if only with near-vanishing intensity when the truth is unimportant enough. Take again the synchronic goal:

G Being such that $(\forall x)(\text{One would now believe } x \text{ if and only if } x \text{ were true})$.

If the scope of the propositional variable here includes the multitude of trivial truths of vanishingly small interest and importance to us, this will presumably induce a correspondingly diminished desire that we satisfy *this* goal, as opposed to the goal that, for all *important* truths P , one would believe P iff P were true. And consider now the implication of this for the account of epistemic aptness through *pursuit of synchronic goal G*. If G is insignificant, the means to it cannot derive high epistemic status thereby. But the epistemic rationality and aptness of a belief in a triviality is *not* proportional to how well that belief furthers our goal G . The trivial truth may be one we only negligibly desire *to believe if and only if it were true*. Irrespective of that, however, it may be epistemically rational to the highest degree. The problem here is that the way in which the truth goal bears on our retail believing is wildly out of step with the degrees of epistemic justification of our unimportant beliefs. Accordingly, the epistemic normative status of *these* beliefs is not plausibly derivable from our interest in believing truths, or from any standing motive towards the truth.

Perhaps we should weaken goal G to G' : $(\forall x)(\text{One would believe } x \text{ only if } x \text{ were true})$. This would be what elsewhere I call *safety* and defend as preferable to its

contraposed *sensitivity*. A safe belief is thus one that you would have only if it were true, whereas a sensitive belief is one that you would *not* have if it were *not* true. So there is, I believe, a lot to be said for requiring safety of any belief candidate for the title of knowledge. Actually, the true requirement will have to be somewhat more complex, since a belief might be unsafe because overdetermined, and yet amount to knowledge.³

Regardless of how the safety goal is to be delineated, a sort of problem may remain. If the objective is to explain the epistemic rationality relevant to whether a belief amounts to knowledge, and if the goal-theoretic strategy is that of understanding such epistemic rationality as a variety of means-ends rationality, then it will be important that the goal be one that potential knowers in fact have and that it be plausible that the positive normative status of our beliefs be explicable through their promoting or being thought to promote the goal in question. But how does our hosting a belief promote the goal of having safe beliefs, goal G' above? It is quite obscure how one promotes such a goal by having any particular belief. Whether we have or do *not* have the belief seems irrelevant to whether we satisfy that goal with respect to the proposition believed. That is to say, whether or not one is such that $(B(p) \rightarrow p)$ seems independent of whether one does or does not actually believe p. Indeed *not* believing that p would seem *less* risky with regard to making sure one does not fall short of one's goal.

In any case, there is now this question: Why think of the epistemically normative status that turns a true belief into knowledge in terms of a *goal*? Why not just say that the belief needs to be safe in order to be knowledge, while making no commitment on whether safety is or is not anyone's goal?

II

Perhaps truth has a role to play *not* as a goal or as a component of a goal but more plausibly as a value in terms of which we can assess beliefs, whether anyone does or should have a corresponding goal or not. We come thus to the value that beliefs have in

³ This point is due to Juan Comesaña. My "Reply to Critics" in *Philosophical Issues* 10 (2000) contains a further reason why a belief might amount to knowledge despite being unsafe.

virtue of being true. And we suppose, for the sake of argument, that truth is the only distinctively cognitive or intellectual intrinsic value or, at least, the only such *fundamental* value. If so, then cognitive methods, processes, faculties, virtues, etc., will have value only derivatively, perhaps in virtue of their efficacy in yielding beliefs that are true. This approach to epistemology is distinctively “reliabilist.” And now, it is argued, we may see how poorly such reliabilism fits our intuitive conviction that knowledge and epistemically rational true belief are more valuable than mere true belief. In brief the argument is as follows.

The Antirealist Argument⁴

1. To believe that p correctly *and* with epistemic rationality is more valuable than merely to believe that p and be right.
2. The additional value of epistemically rational belief over mere true belief would have to derive from the value imported by the belief’s additional property of being thus rational.

⁴ The issues of epistemic normativity involved in this argument are discussed in a growing literature that includes the following, all of which I have found helpful and suggestive, as will be clear to those in the know. (Of course I would not have written this paper had I not been left with a question or two.)

Ward E. Jones, “Why Do We Value Knowledge?” American Philosophical Quarterly 34 (1997): 423-39.

Jonathan.L. Kvanvig, “Why Should Inquiring Minds Want to Know? *Meno* Problems and Epistemological Axiology,” The Monist 81 (1998): 426-51.

Linda Zagzebski, “From Reliability to Virtue Epistemology,” in G. Axtell, ed., Knowledge, Belief, and Character (Lanham, MD: Rowman & Littlefield, 2000).

Marian David, “Truth as the Epistemic Goal,” forthcoming in M. Steup, ed., Knowledge, Truth, and Duty (Oxford: Oxford University Press, 2000).

Michael DePaul, “Value Monism in Epistemology,” forthcoming in M. Steup, ed., Knowledge, Truth, and Duty (Oxford: Oxford University Press, 2000).

Wayne Riggs, “Reliability and the Value of Knowledge,” forthcoming in Philosophy and Phenomenological Research.

In the present paper I develop an approach sketched in “Beyond Skepticism, to the Best of Our Knowledge,” Mind (1988) and in “Reflective Knowledge in the Best Circles,” Journal of Philosophy (1997), reprinted in Steup, *op.cit.* John Greco also treats related issues in his paper for this conference, and I agree with a lot in his contextualist approach.

3. According to reliabilist accounts of epistemic rationality, a belief is epistemically rational through deriving from a method, or process, or faculty, or virtue that is reliable, one that generally yields beliefs that are true.
4. But in that case how can being yielded by such a source add any further value to a belief over and above the value that it has simply in virtue of being true? How can a true belief obtain further value, beyond the value of its truth, by deriving from such a source, when the whole point of using the source is to get beliefs that are true? This would be as absurd as a hedonist supposing that pleasure from a reliable pleasure-source would be better than that pleasure of the same intensity, duration, etc., derived from an unreliable pleasure-source.

With this argument we focus on epistemic rationality, at most one component of epistemic aptness, of what a belief needs in order to qualify as knowledge. I will not here try to relate epistemic rationality more specifically to epistemic aptness. However these are related, since truth is by hypothesis the only fundamental epistemic value, the value of epistemic rationality must itself be explained in terms of truth.⁵

In considering our argument let us first reflect on some varieties of value. One may distinguish first between two sorts of value: the intrinsic and the instrumental. Let us assume monistic hedonism, and consider events Y and Z, each an instance of pleasure. Suppose event Y also brings about much future pleasure, while Z does not. Y is then better than Z, even if it is no better *intrinsically*. Moreover, an event X may not be an instance of pleasure, and hence *not* good intrinsically, while yet it is still good instrumentally, because of the pleasure it yields. All of this we may appreciate, as good hedonists, from a judicial, spectatorial stance that evaluates how matters stand, past, present, and future.

Take now two situations, or even two worlds, wherein the only evaluatively relevant aspects are as follows, for X, Y, and Z as described above. Both worlds

⁵ My value-monistic assumption is only a working assumption. I doubt that the value of understanding can be reduced to that of truth. I should also recognize that my use of 'rationality' here is very broad, and does not pertain only to the proper operation of reason in any narrow sense. It pertains rather more broadly to the proper operation of one's cognitive systems, skin-inwards, or, better, mind-inwards. So epistemic "adroitness" might better capture my sense.

contain this sequence: X occurs, then Y occurs, then Z occurs. In world W1 each member causes its successor if any. In world W2 no member causes any other member. Is world W1 better than world W2? Not according to hedonism, whose only source of intrinsic value is pleasure, for there is no more pleasure in W1 than in W2. And yet the X of W1, call it X1, is anyhow better than the X of W2, call it X2. X1 is better than X2 since X1 brings into the world the intrinsic value that it entrains by causing Y1 and, indirectly, Z1, whereas X2 entrains nothing. And so, X1 is better than X2, and Y1 is better than Y2. And Z1 is the same in value as Z2. And yet W1 is no better than W2. How can this be?

The explanation: Worlds are evaluated by total intrinsic value, but particular events are not. Particular events are also evaluated by their instrumental value, a sort of value with its own distinctive status. True, it is not a fundamental kind of value, since it involves rather the amount of intrinsic value that an event *causes*. So instrumental value is logically constituted by causation plus intrinsic value. The instrumental value of an event derives from the intrinsic value found in the causal progeny of that event. Nevertheless, events can have a distinctively instrumental value over and above any intrinsic value they may also have. When we assess an event from the judicial stance, we may assess it as intrinsically valuable, and also, separately, as instrumentally valuable.

An agent A may bring about an event E. The bringing about of E by A may then itself be assessed. This event, call it E', may not have any intrinsic value beyond the intrinsic value contained already in E, but it will have instrumental value proper to the special relation involved in E's happening because of E'. Call this special sort of instrumental value *praxical value*, the sort of instrumental value in actions of bringing about something valuable. Now, for the hedonist, an event of someone's being pleased does contain some measure of intrinsic value. Supposing that someone brings about that pleasure, is there also value in this further event? There is of course no distinctive intrinsic value, no intrinsic value *beyond* that found in the pleasure brought about. But even a monistic hedonist may yet find in that action some degree of praxical value.

A world does not enhance its total value by containing not only intrinsically valuable but also instrumentally valuable states. Praxical value is in this respect just like any other variety of instrumental value. But, again, from this it does not follow that a

particular event with praxical value is not itself valuable through the praxical value that it contains. Here again praxical value is like instrumental value in general. Instrumentally valuable events do have their proper value, their own sort of *instrumental* value.

III

Take a case of someone's knowing something in particular. *Do* we attribute to such knowledge any value over and above whatever value it has through being a true belief? When a thing has value it has it in respect of having a certain property or satisfying a certain condition. More precisely, then, our question is this: Does a bit of knowledge have value in a respect other than being a true belief? It would seem so, but how could you possibly explain this if you thought that any such additional value must derive from the belief's manifesting an intellectual virtue, understood as a psychological mechanism that would deliver a high enough preponderance of true beliefs (over false ones), at least in normal circumstances. This is hard to see as a respect in which a true belief could then be enhanced, any more than espresso itself is enhanced simply through the reliability of its source.

If persuaded that knowledge must have some value beyond that of its constitutive true belief, therefore, one may well take the Antirealist argument to refute the following sort of virtue epistemology:

- VE
- (i) a belief's epistemic worth is constituted at least in important part through its deriving appropriately from an intellectual virtue, and
 - (ii) what makes a feature of a subject's psychology an "intellectual virtue" is the reliable tendency of that feature to give rise to true beliefs on the part of that subject.

Is VE refuted by the value problem?

Within the sport of archery we aim to hit the bull's-eye, an end intrinsic to that sort of activity. When engaged in the activity, don't we also prefer to hit the bull's-eye by means of skill and not just by luck? A gust of wind might come along and guide our arrow to the bull's-eye, but this will not be as sweet a hit as one unaided by the lucky

gust. Of course a hit that through skill compensates for the wind might be sweeter yet. So I see nothing unacceptable in a notion of a good, skillful shot that goes beyond that of a mere winning or accurate shot. A winning, accurate shot may have been just lucky and not at all skillful, and not in that sense a good shot. In archery we want accurate, winning shots, but we also want shots that are good and skillful. Are the goodness and skill that we want in our shots qualities that we want merely as means? Maybe so, but it seems unlikely. We would not be fully satisfied even with many accurate, winning shots, if they all derived from sheer luck, and manifested no skill, despite our gaining *some* satisfaction through hitting the mark (not to speak of prizes, fame, etc.).

Whether or not we want such goodness and skill only as means, anyhow, a perfectly understandable concept of a good, skillful shot includes *both* hitting the mark *and* doing so through skill appropriate to the circumstances. Can there be any doubt that we have such a concept concerning archery? Surely we do, along with many analogous concepts in other sports, *mutatis mutandis*. What precludes our conceiving of knowledge in a similar way, as a desideratum that includes an intrinsic success component, a hitting of the mark of truth, along with *how* one accomplishes that, how one succeeds in hitting the mark of truth? On this conception, knowledge is not just hitting the mark but hitting the mark somehow through means proper and skillful enough. There seems nothing “incoherent” in any pejorative sense in such a desideratum of “knowledge,” and there are plenty of analogous desiderata throughout the wide gamut of human endeavors.

To recognize that, moreover, may not require us to think that only some additional *intrinsic* value could account for the value in the skillful shot over and above the mere hitting of the bull’s-eye, nor need the additional value be fundamental. The further value might rather be just *praxical* value or the like. Can VE deal with the value problem by appealing thus to the praxical value of hitting the mark of truth through intellectual skill?

IV

In a *very* weak sense even a puppet “does” something under the control of the puppeteer, and even to stumble across a stage unintentionally is to “do” something.

These are cases of “behavior” or even “acts” in correspondingly weak senses. Still a puppet’s performance can be assessed. A puppet can be said to perform well or not, depending for example on whether its hinges are rusty and tend to stick. And the movements of the stumbling ballerina might be, as mere motions, indistinguishable from a lovely pas, though less admirable nonetheless.

Greater independence is displayed by a temperature control system consisting of a thermostat with two triggers, one for a heater and one for a cooler. The system normally keeps the temperature in a certain space within certain bounds. If it gets too hot, then the system triggers the cooler, if it gets too cool, it triggers the heater, and if the temperature is just right, then it idles. What makes it a good system for that space, moreover, is that it *would* perform thus in normal conditions. It is not enough that it *does* perform thus. That the system would perform thus relative to that space is due, finally, to two factors: (a) to its internal constitution and character, and (b) to its relation to the relevant space. In virtue of being stably thus constituted and related, therefore, the system is, at least in a minimal sense, a properly operative system of temperature control for that space.

If it is a sheer accident that it is thus constituted and thus related, then the system falls short in respect of how properly operative it is for that space. At least it falls short in a certain stronger sense of what is required in such a system. A properly operative temperature-control system for a certain space over a certain interval is not one that *accidentally* remains so constituted and related that it would keep the temperature in that space within the desired bounds.

We can of course assess the system independently of its relation to the space. We might naturally assess how well it *would* control the temperature of such spaces if suitably related to them (in a way perhaps in which one can standardly make such systems be related to such spaces). One can of course then assess such a system independently of whether it then happens to be appropriately installed. It might be sitting in a display room in a store. The evaluation would then focus on whether it is so constituted that, if also suitably installed, it would reliably control the temperature of that space.

Whether such a system performs well relative to a certain space would then go beyond whether by virtue of that performance it is or is not contributing causally to keeping the temperature within proper bounds. A good system might perform in such a way that it does so contribute, *not* because it is working right, however, but only because, although it does not then work right, luck enables it to cause the right outcome anyhow. Thus the system may suffer a glitch, while yet, coincidentally, an insect happens to alight on a crucial component of its internal mechanism so that the system does trigger the cooler as it should if it is to keep the temperature within the proper bounds. Only because of that bit of luck does the system then contribute causally to keeping the temperature within the proper bounds. Had it not been for the insect, then, it would not have triggered the cooler. So it would be false to say that it “worked right” on that occasion, when it just suffers a glitch. It is a good enough system nevertheless, since it does work right in the great majority of circumstances where it is normally called on to operate.⁶

This example shows that for a system to work right or perform well, in ways that are creditable, more is required than just (a) that it is be good system (in the relevant respects), and (b) that it then contribute causally to the desired outcome. For it may contribute causally to that outcome only through some fluke, in which case it then contributes despite *not* working right or performing well.

A system works right or performs well on a particular occasion, then, only if it unflukily enters a state that would lead in the relevant circumstances to the desired outcome. Accordingly, what the system does in entering that state, unaided by luck, must be sufficient to produce the desired outcome, given perhaps its normal relation to its relevant space. In our example the system did enter such a state, but only because of the insect in machina.

V

⁶ It might be replied that the system did work right, ... *with the help of the insect*. And I am in some linguistic sympathy with this reply. Perhaps, I am willing to grant, ‘working right’ is at least ambiguous, and in one sense it does permit this reply. In any case, there would presumably remain the sense in which the system itself does not really work right, which is tantamount to its not working *well* on that occasion, and ‘working well’ lends itself less well to the present reply, as it seems less subject to the ambiguity that affects ‘working right’.

An artifact like our temperature control system that “does” things, that “works,” might be evaluated variously, along with its performances. We might evaluate it by reference to how well it serves those it is expected to serve in certain characteristic ways. Or we might evaluate it independently of how well installed it is, if installed at all. Or we might evaluate its operation on a particular occasion: does it work right or perform well on that occasion? So there is the “agent” in a broad sense that includes mechanical agents of various degrees of sophistication, with various ranges of intended activity. There is the “performance” of that agent on a particular occasion. And sometimes there is also a performance-distinct situation or object or quantity of stuff that such an agent brings about or produces through its performance, a performance-distinct result that might also be performance-transcendent.

None of our three evaluations of aspects of such a situation, wherein an agent performs, uniquely determines any of the others. An agent could be a fine agent and perform poorly on a particular occasion, which in turn is compatible with the performance-distinct product being of high quality or of low quality or of any quality in between. Or the agent could be a mediocre or worse agent and yet perform well on a given occasion. It is even possible that a poor performance by a poor agent may lead to an excellent performance-distinct outcome. So the three dimensions of evaluation seem largely independent of each other—but not entirely, or so I now argue.

Performances relate in one direction to the agents involved, and in another direction to their performance-distinct products, if any. So they might be evaluated with a view towards one direction, or towards the other. An agent might be nearly incompetent and yet perform most effectively on a particular occasion. This evaluates the performance in the light of its wonderful outcome. Someone with a barely competent tennis serve may blast an ace past his opponent at 130 mph. This is a most effective serve given its outcome: a ball streaking past the receiver untouched, having bounced within the service court. But from another point of view it may not have been so positively evaluable after all. If the player is a rank beginner, for example, one most unlikely to reproduce that performance or anything close to it, then one may reasonably withhold one’s encomium. It was still a wonderfully *effective* serve, but hardly a *skillful* one. Performances are in this way double-faced. So the evaluation of a performance

seems not after all independent of the evaluation of the other two components of the performance situation. Either the evaluation is agent-involving or it is outcome-involving (or both). Performances that are creditable must be attributable to the agent's skills and virtues, and thus attributable to the agent himself.

VI

In evaluating the Antirealist Argument it will help to have in view the categories of praxis, of human doings and actions. To get something done you do not need much sophistication. Water flows downhill, for example, supermarket doors open when people approach, your knee jerks under the doctor's mallet, and so on. Distinctively human action is on a higher plane. A rational agent's action is controlled and informed by reason. At a minimum one must know what one is doing and must do it for a reason. Bees dance, it is true, perhaps guided not only by instincts but also by "reasons," unlike puppets. Higher up the animal kingdom, in any case, and well before we reach humanity, much behavior is less and less plausibly explained by appeal to mere instinct. Differences of degree are still differences, however, and the rational animal stands head and shoulders above the rest.

Three sorts of agency. The dimension of agency divides into at least three divisions, corresponding to at least three ways of \emptyset 'ing. First, there is *plain* \emptyset 'ing, however autonomously or informedly, or even attributively. Second, there is \emptyset 'ing intentionally, perhaps deliberately. Third, there is \emptyset 'ing *attributably*, in a way that makes one's \emptyset 'ing attributable to oneself as agent. Just as it is fallacious to infer that an agent \emptyset 's intentionally from the fact that he \emptyset 's, since he might \emptyset unintentionally, so it is a fallacy to infer that an agent \emptyset 's attributably from the fact that he \emptyset 's, since he might \emptyset unattributably, as when someone *falls*, having been pushed off a roof. All intentional \emptyset 'ings are attributable, but the converse is false; and all attributable \emptyset 'ings are plain \emptyset 'ings, but again the converse is false.

Three forms of evaluation. With regard to instruments, tools, mechanisms, and useful artifacts, methods, and procedures in general, there are three interestingly different forms of evaluation, positive or negative, whether the evaluation takes the form of approval, favoring, or admiring, or the form of disapproval, disfavoring, or

deploring. A useful cultural device is normally meant to help secure goods that we value independently of the device. Thus we value conveyance to one's destination, ambient temperature within certain bounds, savory and nourishing food, etc. And we also value devices whose normal operation will enable us to secure those goods, and also particular instances where the operation of the device secures one of its characteristic goods.

We favor and approve of good performance in our devices. We admire and even "praise" such performance. On the flip side, we disfavor and disapprove of malfunction, and deplore poor performance, and may even "blame" it on the device. Such evaluations of performance, whether positive or negative, go beyond the evaluation of goods produced by the performance, whether it be conveyance to one's destination, ambient temperature within certain bounds, savory and nutritious food, etc. And they also go beyond evaluation of the artifact and of its general reliability. The evaluation of a particular performance is distinct from the evaluation of the artifact that then performs and of any performance-transcendent product of the performance. A first-rate artifact may yield an excellent product despite the very low quality of its performance itself on that occasion.⁷

Why do we evaluate not only devices and their products, but also, separately, their performances? Well, why do we evaluate not only intrinsic value but also extrinsic value? Presumably we have concepts of instrumental value because it is useful for us to keep track of the levers of useful power. We bend nature to our ends, and in doing so we rely on what works, on what leads causally to our desiderata. Thus the importance of suitable concepts that help us keep track of what does work, of what has value through its causal powers. These are often states we can bring about more directly, whereby we secure more remote effects, as when we switch on the light by flipping the switch. But an extrinsically evaluable state need *not* be such a potential instrument relative to human capacities. A hurricane can be awful even if uncontrollable. The more general concept is the concept of what brings about value or disvalue; it is the more general concept of the extrinsically good or bad. Nevertheless, such a concept seems

clearly important to agents whose wills must work indirectly in securing outcomes desired for their intrinsic worth and in avoiding those intrinsically unworthy. And it also seems important to those who need to adjust their conduct in the light of perceived danger, regardless of our ability to control that sort of situation. Thus the hurricane; we can at least control our relationship to it.

Tools, instruments, mechanisms, and other artifacts and devices, draw our interest primarily for the goods secured through their use. Efficiently smooth operation may of course be admired in its own right irrespective of its utilitarian implications. For the most part, nonetheless, what we care about in our artifacts is that they serve us well by helping produce the goods that we want from them.

Derivatively from that, we evaluate also the artifacts themselves in respect of how reliably they operate in the normal circumstances of their operation. We need to keep track of the reliability of our thermostats, cars, airplanes, etc., so we need concepts, including evaluative concepts, that enable us to discriminate the reliable from the unreliable. And we evaluate the *performances* of our artifacts, using similar categories of evaluation. What is the *point* of such evaluation? If we already know (a) that the performance-transcendent product of the performance has a certain value, and (b) that the performing artifact also is reliable up to a certain level, why then are we *also* interested in (c) how worthy the particular performance is?

It is hard to see what interest there could be in evaluations of artifactual performance except through the implications of such evaluations for assessment of the performing artifacts. Thus we have an important interest in the reliable quality of our artifacts, and from this interest derives rationally our interest in the quality of their particular performances. (Again I am leaving aside any purely aesthetic interest that artifactual performances may acquire.) Artifacts are “agents,” however, only weakly, perhaps in an extended and even metaphorical sense. And this is of a piece with our treating them as mere means (except when we treat them as objects of aesthetic appreciation). Correlatively, our approval, admiration, and even praise for their

⁷ And things can come unravelled in other ways too. Thus an excellent performance may have an unfortunate outcome due to unfavorable circumstances.

performances is also qualified by the standing of the performers as mere tools at the service of our ends.

VII

Wise action. Suppose your raft glides downstream and comes to a fork. Down the right effluent there's treasure, down the left effluent only mud. Knowing this, and having control of the rudder, you take the right effluent and reach your reward. What you do and your attainment are then attributable to you, and properly admirable and praiseworthy. The reward is something you win through your own well-directed rational effort.

Consider now some ways one might fall short of that:

- (a) The raft might be completely beyond one's control, either because someone else controls its rudder and disregards your preferences, or because it drifts rudderless.
- (b) One might not know that one is going down the right effluent, or that it is better to take that direction.

If either (a) or (b) is true of you, then even if you *do* go down the right effluent and reach the prize, this will be something that *happens* to you, by luck; it will not be something properly attributable to you as your rational doing, as something properly admirable in you, or as something properly deserving of praise.

Again, in a *very* broad sense you *do* something when you "go down the right effluent tied down and blindfolded." You do something at least as does water when it flows, as does the knee when it jerks. Take an arbitrary "doing" of yours, in this very general sense. What conditions must such a doing satisfy in order to qualify as a proper subject not only of admiration (as one may admire the swelling flow of Niagara) but as a proper basis for praise or blame, credit or discredit? One condition would require that it be autonomous enough, another that it be sufficiently well informed.

What is true of our doings generally is true of our believings in particular. Belief, too, may be found up and down the evolutionary scale, and even below. A door may

“think” somebody is approaching when a garbage can blows by. A dog may think it’s about to be fed when it hears a clatter in the kitchen. Man is rational when his belief is controlled and informed by reason (and “adroit” through his properly operating cognitive systems). We may believe in the weakest sense, however, as when a belief is instilled through subliminal suggestion or through hypnosis or brainwashing. Such a belief is insufficiently derived from the exercise of the distinctively intellectual capacities and abilities, the faculties, cognitive methods, and intellectual virtues of the subject (irrespective of whether its adoption counts as *voluntary*). The believing is hence not attributable to the subject, not even in the way in which the circulation of the blood is attributable to the subject’s heart and thereby, indirectly, to the subject as well. You may believe something, again, in a way that does not derive from the exercise of your intellectual excellences, but only from some external source not appropriately under your cognitive control. If so, then the believing in question may not be properly attributable to you as your doing. It may be something you *do* only as weakly as does the puppet dance when the puppeteer makes it do so.

VIII

Even when one does attributably bring about one’s belief, one’s believing something in particular, it remains to be seen whether its being a *true* believing is also attributable to oneself as one’s own doing. The following is, again, fallacious:

1. Attributably to S as S’s doing, S \emptyset ’s.
2. S \emptyset ’s in way W.
3. Therefore, attributably to S as S’s doing, S \emptyset ’s in way W.

Consider:

- 1a. Attributably to your heart as its doing, it pumps blood through your body.
- 2a. Your heart pumps blood in this building (pointing to the building where you are).
- 3a. Therefore, attributably to your heart as its doing, it pumps blood in this building.

Your heart's pumping blood in this building is perhaps your doing, since for one thing you could easily have been elsewhere, but that the pumping takes place in this building is not attributable to your heart as *its* doing.

Analogously, the following would also be fallacious.

- 1b. Attributably to you as your doing, you believe that p.
- 2b. You believe that p correctly (with truth).
- 3b. Therefore, attributably to you as your doing, you believe that p correctly.

So in order for *correct* belief to be attributable to you as your doing, the being true of your believing must derive sufficiently from "yourself," which involves its deriving from constitutive features of your cognitive character, and of your psychology more generally.

If truth has its own cognitive or intellectual value, then bringing about one's believing truly will have its corresponding praxical value, a distinctive sort of instrumental value. Compatibly with this, truth may still have a special role in explaining the normativity of belief. For the hedonist, similarly, pleasure has a special role in explaining the normativity of action, even if there are many things with value besides instances of pleasure. Eating savory food will have value instrumentally by promoting pleasure, for example, and the bringing about of pleasure will have its own distinctive value, different from the intrinsic value of the pleasure brought about, but value nonetheless, praxical value.

Does the bringing about of pleasure have value over and above the value of the contained pleasure? Well, it does have a different sort of value, one distinct from the value of the pleasure brought about. So a world where that pleasure is present uncaused will have the same intrinsic value as this one, but it will be missing something present here, which does here have value of a sort, praxical value. This is rather like the comparison between the world where the X-Y-Z sequence occurs unaided by any causation, and hence with no instrumental value in the X or the Y components, by comparison with the world where Y causes Z and X causes Y, wherein there is the same intrinsic value present in the three items, but wherein also (a) it's a good thing X

happens, *not* because of any intrinsic value of its own but because of the intrinsic value that it yields by causing Y and Z, and (b) it's a good thing Y happens, not only because of its own intrinsic value but also because of the intrinsic value that it yields by causing Z

Similarly, in the world where an agent brings about some pleasure, there is not only the intrinsic value of the pleasure but also the distinctive value of the agent's action. We can say about that action that it's good that it is done, at least in the respect that it brings about some pleasure, which is intrinsically good. So there is this practically good action in the world in addition to the intrinsically good pleasure that it brings about.

Consider now a case where a true belief, a *true* believing is attributable to you as your doing. We may now say that, besides the epistemic good in that true belief, there is further the practical good in your action of bringing it about. And this arguably involves your exercise of excellences constitutive of your cognitive character.

That is, it seems to me, a way in which truth can have a distinctively important and fundamental place in explaining epistemic normativity, compatibly with knowledge having epistemic worth over and above the worth of mere true belief. We can see the good proper to an epistemic action creditable to the agent, who brings about that good for himself, and is more than just the recipient of blind epistemic luck.⁸

IX

However, the account of the extra value of knowledge in terms of the practical value that it contains does not go far enough. For this practical value does not explain the fact that we would prefer a life of knowing, where we gain truth through our own intellectual performance, to a life where we are visited with just as much truth but through mere external agency (brainwashing, hypnosis, subliminal suggestion, etc.). This

⁸ In fact the account here of practical value is only a first approximation, perhaps sufficient unto the day. A more adequate account, in any case, would allow the possibility of a performance with practical value that does *not* succeed in securing its characteristic inherent value. Even if some bad luck robs it of its expectable fruits, an action may still be a wonderful performance, and properly admirable, and correspondingly valuable, *practically* valuable in our richer sense. (Delineating that sense should be within reach, and what follows is one attempt.)

might be the work of a less malevolent evil demon, who allows a world out there pretty much as we believe it to be, but one that fits our beliefs only through happenstance, the happenstance that the demon has deigned to give us just those beliefs although he might more easily have given us beliefs dissonant from our external reality.⁹

If we prefer a life in which we gain our truths through our own performances, then the value of our apt performances cannot be mere praxical value. For if it were merely praxical value, then the value of our performances would derive entirely from their causing the intrinsic value resident in the true believings that they would bring about. In that case, and if true believing is the only intrinsically valuable epistemic good, then two worlds containing the same true believings could hardly differ in overall value, regardless of the fact that in one of them there is a lot more praxical value. Compare the case of extrinsic value more generally, and the two X-Y-Z worlds above.

So if we rationally prefer a world in which our true beliefs derive from our own cognitive performances to one with the same true beliefs, now courtesy of the less malevolent demon, then there must be some further value involved in the first world not exhausted merely by the praxical value that it contains. What could this further value be?

When Aristotle speaks of the “chief good” as activity which goes beyond the state of mind that produces it since “the state of mind may exist without producing any good result” it seems clear that in his view performances creditable to an agent as their own are the components of eudaimonia, of human good or faring well, which “turns out to be activity of soul in accordance with virtue.” In purely theoretical activity, moreover, truth and falsity are the good and bad state respectively, and the work of everything intellectual.

According to the Aristotelian view, then, passive reception of truth is not enough to count as human good, or at least not as the chief human good. Our preference is not just the presence of truth, then, however it may have arrived there. We prefer truth whose presence is the work of our intellect, truth that derives from our intellectual performance. We do not want just truth that is given to us by happenstance, or by

⁹ Nor does the account explain how the virtuous bringing about of a true belief is better than the accidental bringing about of that belief even if the two bringings about are otherwise the same to the greatest possible extent.

some alien agency, where we are given a belief that hits the mark of truth *not* through our own performance, but in a way that represents no accomplishment creditable to us.¹⁰

We have reached the following result. Truth-connected epistemology might grant the value of truth, of *true believing*, might grant its intrinsic value, while allowing also the praxical extrinsic value of one's attributably hitting the mark of truth. This praxical extrinsic value would reside in such attributable intellectual deeds. But in addition to the extrinsic praxical value, we seem plausibly committed to the *intrinsic* value of such intellectual deeds. So the grasping of the truth central to truth-connected reliabilist epistemology is not just the truth that may be visited upon our beliefs by happenstance or external agency. We desire rather truth gained through our own performance, and this seems a reflectively defensible desire for a good preferable not just extrinsically but intrinsically. What we prefer is the deed of true believing, where not only the believing but also its truth is attributable to the agent as his or her own doing.

Does this adequately account in reliabilist terms for the value of knowledge over and above its contained true belief? Is the additional value simply the value contained in the attributable, creditable attaining of the truth, as opposed to the mere presence of truth (which might conceivably derive from happenstance or external agency)? The foregoing considerations go quite far, but *not* all the way to the required full account of epistemic value within reliabilist, truth-connected epistemology. At least one further step is needed and that is the aim of our next section.

X

Compare two evil demon victims. The first victim takes in quite fully and flawlessly the import of her sense experience and other states of consciousness, to an extent rarely matched by any human, and then reasons therefrom with equal prowess to conclusions

¹⁰ When I presented these ideas at Notre Dame, Alvin Plantinga wondered what would be so bad about being the beneficiary of Divine revelation, where there are no special faculties, really, that set one apart; where one is just visited by the overpowering light of the revealed truth. In response it seemed to me that even if, with Aristotle, one finds the de facto chief human good in active virtuous attainments of one's own, this need not prevent one from granting that there may be other ways to the truth that might be just as desirable and even admirable. It seems to me that much of our epistemology and epistemic value theory could be isolated from such issues of rational theology.

beyond the reach of most people, and retains her results in memory well beyond the normal. The other victim is on the contrary extensively handicapped in her cognitive faculties and performs with singular ineptness. Clearly one of these victims is better off than the other; you would prefer to be and perform like the first and unlike the second. However, neither one attains truth at all, not even as a doing, through being visited with truth; much less does either one attain truth as a deed, by hitting the mark of truth through the excellence of their performance. So the epistemic value of the intellectual conduct of the first victim, the value that lifts her performance over that of the second victim, is not to be explained in the terms of our earlier account. Neither subject hits the mark of truth at all, whether attributably and creditably or not. So how can one of them still attain more value than the other? What sort of value can this be?

Recall the temperature-control device, with the two triggers. Suppose it is taken off the shelf in the display room for a demonstration, and a situation is simulated wherein it should activate the cooling trigger, and then a second situation is simulated wherein it should activate the warming trigger. In such a test the device might either perform well or not. But the quality of its performance is not to be assessed through how well it actually brings about the goods that it is meant to bring about in its normal operation. For in the display room it brings about neither the cooling nor the heating of any space. And yet we can and do assess the quality (and in a sense the “value”) of its performance. What we are doing is quite obvious: we are assessing whether it performs in ways that would enable it to bring about the expected goods once it was properly installed, i.e., properly related to the target of its operation. We might call this sort of value “performance value.” The performance value of a performance is the degree of positive or negative quality attained by that operation, measured by how well the performance enables the “agent” to operate, by entering various states in various circumstances, so as to be such that, when suitably installed, it would in fact bring about the expected goods in its target (where of course the “agent” and “target” might be the same).

It does not require an imaginative leap to conceive of our cognitive systems as devices that operate normally with the expected result: truths of certain sorts acquired by the host organism. There are various ways of conceptualizing this, but one way

might include *the visual system, the auditory system*, and so on. Alternatively, we might have *the brain-including nervous system, together with sense organs*. Alternatively, we might have *the animal, or the human being*. In any case, there would be the system or organism on one side, and the normal environments in which it operates on the other. And we can evaluate the performances of the organism independently of its proper emplacement in a suitable environment. This would be similar to what we do in evaluating the performance of the temperature-control device in the display room. Consider then the deliverings of our cognitive systems, of whatever level of complexity we pick, including the top, total-human, level. Such deliverings can be assessed for performance value, through assessing how well the performance would enable the system to deliver the expected goods if it were “properly installed” in a suitable environment.

Recall the greater epistemic value, the higher epistemic quality, found in the performance of the first of our two victims of the less malevolent demon. We may now say that this higher value is performance value. It is like the higher value of the glitch-free performance of the temperature control device in the display room under simulation. If this is correct, then we have a way to understand the value of the epistemic justification that we find in the beliefs of the properly “perceiving” and reasoning victim of the evil demon. It is performance value, and what is good about this performance value is still to be understood in a truth-connected, reliabilist way. What is good about that performance value cannot be understood independently of the fundamental value of true believing, and especially of true believing that hits the mark of truth attributable to the agent. For *this* is the good that the relevant system is expected to deliver through its operation when “properly installed” in a suitable environment, and the good that may thus be credited to the organism as a whole, in virtue of the proper operation of its cognitive architecture.

Does that sufficiently identify the sorts of epistemic values that an adequate epistemology should be able to explain? We have identified: (a) the value of bare true believing (since we do prefer to be given truth rather than falsehood, even when it comes through happenstance or external agency); (b) the praxical, *extrinsic* value of true believing where the agent brings about the belief, and perhaps even hits the truth

as his own doing; (c) the praxical, *intrinsic* value of true believing where the agent hits the mark of truth as his own attributable deed, one which is hence creditable to the agent as his own doing; and (d) the performance value of a deliverance-induced believing, present even when the belief induced is false, so long as the performance is high on the quality scale for such performances, as measured by how well such performance would provide the expected goods, if the system were properly installed, in a suitable environment. Are we able to account for all our intuitions concerning epistemic evaluation, epistemic quality and value, in terms of these four concepts of epistemic normative or evaluative status?

It might be objected now again that our preference for the life of the first of the two demon victims, the one who “perceives” and reasons properly is not explained exhaustively merely through appeal to the performance value of the believings of that victim. For if it were mere performance value, then we would *not* hold that world and that life to be intrinsically better than the life and world of the other victim. But we do think it to be thus intrinsically better, do we not?

Surely we care about our devices performing well in display rooms not intrinsically (again, leaving aside aesthetic evaluation) but only because that shows them to be devices suitable for delivering the goods. But it is the goods to be delivered that we really care about. Of course the goods to be delivered need not be performance-transcendent. And indeed, on the Aristotelian view, in our intellectual lives the goods to be delivered by our cognitive systems are not performance-transcendent. The “chief” intellectual goods involve attributable truth-attainment, where one does hit the mark of truth through the quality of one’s performance. Nevertheless, one cares about cognitive systems in good working order not for their own sake, but for the truth-attaining performances that they enable. Much less does one care about good performances by cognitive systems “in display rooms” isolated from the environments within which they would enable one attributably to attain the truth. Such good performances are valued presumably only for their implications about the worth of the operative systems, so their value is, it seems to me, partly epistemic; they manifest within our view the worth of the operative systems. But partly it is a distinctive value of its own, even independently of what they enable us to know. Even if there is no-one around to see it,

the good performance by a system is somehow better than its poor performance; and this is presumably at least in part a matter of the more-than-accidental connection between the quality of the performance and the quality of the system. To the extent that the system performs poorly, to that extent is it a lesser system than it might be.

In any case, whether through its epistemic value or through its connection with the worth of the performing system, the value of simulational good performance is, like extrinsic value, not of fundamental, intrinsic import. The world with such good performances is no better epistemically on the whole than the one without them, so long as the two worlds contain all the same intrinsically valuable epistemic goods.

If so, then those who defend the fundamental status of truth or truth attainment at the basis of epistemic value would seem committed to denying that the good performance of the superior victim is of a higher intrinsic order than the poor performance of the other victim. They are different in quality, true enough, those two performances, but the difference is to be explained in terms of performance value, and hence not in terms of intrinsic value. That is how it would seem on the eudaimonistic account, and that is how it seems to me.¹¹

¹¹ Much of our reflection in epistemology seems applicable to ethics, *mutatis mutandis*.