

## Reliabilism and intellectual virtue

Externalism and reliabilism go back at least to the writings of Frank Ramsey early in this century.<sup>1</sup> The generic view has been developed in diverse ways by David Armstrong, Fred Dretske, Alvin Goldman, Robert Nozick, and Marshall Swain.<sup>2</sup>

### A. GENERIC RELIABILISM

Generic reliabilism might be put simply as follows:

S's belief that p at t is justified iff it is the outcome of a process of belief acquisition or retention which is reliable, or leads to a sufficiently high preponderance of true beliefs over false beliefs.

That simple statement of the view is subject to three main problems: the generality problem, the new evil-demon problem, and the meta-incoherence problem (to give it a label). Let us consider these in turn.

*The generality problem* for such reliabilism is that of how to avoid processes which are too specific or too generic. Thus we must avoid a process with only one output ever, or one artificially selected so that if a belief were the output of such a process it would indeed be true; for every true belief is presumably the outcome of some such too-specific processes, so that if such processes are allowed, then every true belief would result from a reliable process and would be justified. But we must also avoid processes which are too generic, such as perception (period), which surely can produce not only justified beliefs but also unjustified ones, even if perception is on the whole a reliable process of belief acquisition for normally circumstanced humans.<sup>3</sup>

1 Frank Ramsey, *The Foundations of Mathematics and Other Logical Essays* (London: Routledge & Kegan Paul, 1931).

2 David Armstrong, *Belief, Truth and Knowledge* (Cambridge University Press, 1973); Fred Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971): 1-22; Alvin Goldman, "What Is Justified Belief?" in George Pappas, ed., *Justification and Knowledge* (Dordrecht: D. Reidel, 1979); Robert Nozick, *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press, 1981), chapter 3; Marshall Swain, *Reasons and Knowledge* (Ithaca, N.Y.: Cornell University Press, 1981).

3 This problem is pointed out by Goldman himself (*op. cit.*, p. 12), and is developed by

## THEORIES OF JUSTIFICATION

*The evil-demon problem* for reliabilism is not Descartes's problem, of course, but it is a relative. What if twins of ours in another possible world were given mental lives just like ours down to the most minute detail of experience or thought, etc., though they were also totally in error about the nature of their surroundings, and their perceptual and inferential processes of belief acquisition accomplished very little except to sink them more and more deeply and systematically into error? Shall we say that we are justified in our beliefs while our twins are not? They are quite wrong in their beliefs, of course, but it seems somehow very implausible to suppose that they are unjustified.<sup>4</sup>

*The meta-incoherence problem* is in a sense a mirror image of the new evil-demon problem, for it postulates not a situation where one is internally justified though externally unreliable, but a situation where one is internally unjustified though externally reliable. More specifically, it supposes that a belief (that the President is in New York) which derives from one's (reliable) clairvoyance is yet *not* justified if either (a) one has a lot of ordinary evidence against it, and none in its favor; or (b) one has a lot of evidence against one's possessing such a power of clairvoyance; or (c) one has good reason to believe that such a power could not be possessed (e.g., it might require the transmission of some influence at a speed greater than that of light); or (d) one has no evidence for or against the general possibility of the power, or of one's having it oneself, nor does one even have any evidence either for or against the proposition that one believes as a result of one's power (that the President is in New York).<sup>5</sup>

### B. GOLDMAN'S RELIABILISMS

How might reliabilism propose to meet the problems specified? We turn first to important work by Goldman, who calls his theory "Historical Reliabilism," and has the following to say about it:

The theory of justified belief proposed here, then, is an *Historical* or *Genetic* theory. It contrasts with the dominant approach to justified belief, an approach that generates what we may call (borrowing a phrase from Robert Nozick) *Current Time-Slice* theories. A *Current Time-Slice* theory makes the justificational status of a belief wholly a function of what is true of the cognizer *at the*

Richard Feldman in "Reliability and Justification," *The Monist* 68 (1985): 159–74.

4 This problem is presented by Keith Lehrer and Stewart Cohen in "Justification, Truth, and Coherence," *Synthese* 55 (1983): 191–207.

5 This sort of problem is developed by Laurence Bonjour in "Externalist Theories of Empirical Knowledge," in *Midwest Studies in Philosophy, Vol. 5: Studies in Epistemology*, ed. P. French et al. (Minneapolis: University of Minnesota Press, 1980).

*time* of the belief. An Historical theory makes the justificational status of a belief depend on its prior history. Since my Historical theory emphasizes the reliability of the belief-generating processes, it may be called *Historical Reliabilism*.<sup>6</sup>

The insights of externalism are important, and Goldman has been perceptive and persistent in his attempts to formulate an appropriate and detailed theory that does them justice. His proposals have stimulated criticism, however, among them the three problems already indicated.

Having appreciated those problems, Goldman in his book<sup>7</sup> moves beyond Historical Reliabilism to a view we might call rule reliabilism, and, in the light of further problems,<sup>8</sup> has made further revisions in the more recent "Strong and Weak Justification." The earlier theory, however, had certain features designed to solve the new evil-demon problem, features absent in the revised theory. Therefore, some other solution is now required, and we do now find a new proposal.

Under the revised approach, we now distinguish between two sorts of justification:

A belief is *strongly justified* if and only if it is well formed, in the sense of being formed by means of a process that is truth-conducive in the possible world in which it is produced, or the like.

A belief is *weakly justified* if and only if it is blameless though ill-formed, in the sense of being produced by an unreliable cognitive process which the believer does not believe to be unreliable, and whose unreliability the believer has no available way of determining.<sup>9</sup>

Notice, however, that it is at best in a *very* weak sense that a subject with a "weakly justified" belief is thereby "blameless." For it is not even precluded that the subject take that belief to be very ill-formed, so long as he is in error about the cognitive process that produces it. That is to say, S might hold B, and believe B to be an output of P, and hold P to be an epistemically unreliable process, while in fact it is not P but the equally unreliable P' that produces B. In this case S's belief B would be weakly justified, so long as S did not believe P' to be unreliable, and had no available means of determining its unreliability. But it seems at best extremely strained to hold S epistemically "blameless" with regard to

6 See Goldman, "What Is Justified Belief?" pp. 13–14.

7 Alvin Goldman, *Epistemology and Cognition* (Cambridge, Mass.: Harvard University Press, 1986); *idem.*, "Strong and Weak Justification," in *Philosophical Perspectives*, Vol. 2: *Epistemology* (1988): 51–71.

8 Some of these are pointed out in my "Beyond Scepticism, to the Best of our Knowledge," *Mind* 97 (1988): 153–88.

9 Goldman, "Strong and Weak Justification," p. 56.

#### THEORIES OF JUSTIFICATION

holding B in such circumstances, where S takes B to derive from a process P so unreliable, let us suppose, as to be epistemically vicious.

The following definition may perhaps give us a closer approach to epistemic blamelessness.

A belief is *weakly justified (in the modified sense)* if and only if it is blameless though ill-formed, in the sense of being produced by an unreliable cognitive process while the believer neither takes it to be thus ill-formed nor has any available way of determining it to be ill-formed.

With these concepts, the Historical Reliabilist now has at least the beginnings of an answer both for the evil-demon problem and for the meta-incoherence problem. About the evil demon's victims, those hapless twins of ours, we can now say that though their beliefs are very ill-formed – and are no knowledge even if by luck they, some of them, happen to be true – still there is a sense in which they are justified, as justified as our corresponding beliefs, which are indistinguishable from theirs so far as concerns only the “insides” of our respective subjectivities. For we may now see their beliefs to be weakly justified, in the modified sense defined above.<sup>10</sup>

About the meta-incoherence cases, moreover, we can similarly argue that, in some of them at least, the unjustified protagonist with the wrong (or lacking) perspective on his own well-formed (clairvoyant) belief can be seen to be indeed unjustified, for he can be seen as subjectively unjustified through lack of an appropriate perspective on his belief: either because he positively takes the belief to be ill-formed, or because he “ought” to take it to be ill-formed given his total picture of things, and given the cognitive processes available to him.

Consider now the following definition:

A belief is *meta-justified* if and only if the believer does place it in appropriate perspective, at least in the minimal sense that the believer neither takes it to be ill-formed nor has any available way of determining it to be ill-formed.

Then any belief that is weakly justified (again, sticking to the unmodified sense) will be meta-justified, but there can be meta-justified beliefs which are not weakly justified. Moreover, no strongly justified belief will be weakly justified, but a strongly justified belief can be meta-justified. Indeed one would wish one's beliefs to be not only strongly justified

<sup>10</sup> I will use the modified sense in what follows because it seems clearly better as an approach to blamelessness; but the substance of the critique to follow would apply also to the unmodified sense of weakly justified belief.

but also meta-justified. And what one shares with the victim of the evil demon is of course not weak justification. For if, as we suppose, our own beliefs are strongly justified, then our own beliefs are not weakly justified. What one shares with the evil demon's victim is rather meta-justification. The victim's beliefs and our beliefs are equally meta-justified.

Does such meta-justification – embedded thus in weak justification – enable answers both for the new evil-demon problem and for the problem of meta-incoherence? Does the victim of the evil demon share with us meta-justification, unlike the meta-incoherent? The notion of weak justification does seem useful as far as it goes, as is the allied notion of meta-justification, but we need to go a bit deeper,<sup>11</sup> which may be seen as follows.

### C. GOING DEEPER

Beliefs are states of a subject, which need not be occurrent or conscious, but may be retained even by someone asleep or unconscious, and may also be acquired unconsciously and undeliberately, as are acquired our initial beliefs, presumably, whether innate or not, especially if deliberation takes time. Consider now a normal human with an ordinary set of beliefs normally acquired through sensory experience from ordinary interaction with a surrounding physical world. And suppose a victim in whom evil demons (perhaps infinitely many) implant beliefs in the following way. The demons cast dice, or use some other more complex randomizer, and choose which beliefs to implant at random and in ignorance of what the other demons are doing. Yet, by amazing coincidence, the victim's total set of beliefs is identical to that of our normal human. Now let's suppose that the victim has a beautifully coherent and comprehensive set of beliefs, complete with an epistemic perspective on his object-level beliefs. We may suppose that the victim has meta-justification for his object-level beliefs (e.g., for his belief that there is a fire before him at the moment), at least in the minimal sense defined above: he does not believe such beliefs to derive from unreliable processes, nor has he any available means of determining that they do. Indeed, we may suppose that he has an even stronger form of meta-justification, as follows:

S has meta-justification, in the stronger sense, for believing that p iff  
 (a) S has weaker meta-justification for so believing, and (b) S has

<sup>11</sup> Though, actually, it is not really clear how these notions will deal with part (d) of the problem of meta-incoherence: cf. Goldman, *Epistemology and Cognition*, pp. 111–12.

#### THEORIES OF JUSTIFICATION

meta-beliefs which positively attribute his object beliefs in every case to some faculty or virtue for arriving at such beliefs in such circumstances, and further meta-beliefs which explain how such a faculty or virtue was acquired, and how such a faculty or virtue, thus acquired, is bound to be reliable in the circumstances as he views them at the time.

And the victim might even be supposed to have a similar meta-meta-perspective, and a similar meta-meta-meta-perspective, and so on, for many more levels of ascent than any human would normally climb. So everything would be brilliantly in order as far as such meta-reasoning is concerned, meta-reasoning supposed flawlessly coherent and comprehensive. Would it follow that the victim was internally and subjectively justified in every reasonable sense or respect? Not necessarily, or so I will now try to show.

Suppose the victim has much sensory experience, but that all of this experience is wildly at odds with his beliefs. Thus he believes he has a splitting headache, but he has no headache at all; he believes he has a cubical piece of black coal before him, while his visual experience is as if he had a white and round snowball before him. And so on. Surely there is then something internally and subjectively wrong with this victim, something "epistemically blameworthy." This despite his beliefs being weakly justified, in the sense defined by Goldman, and despite his beliefs being meta-justified in the weaker and stronger senses indicated above.

Cartesians and internalists (broadly speaking) should find our victim to be quite conceivable. More naturalistic philosophers may well have their doubts, however, about the possibility of a subject whose "experience" and "beliefs" would be so radically divergent. For these there is a different parable. Take our victim to be a human, and suppose that the demon damages the victim's nervous system in such a way that the physical inputs to the system have to pass randomizing gates before the energy transmitted is transformed into any belief. Is there not something internally wrong with this victim as well, even though his beliefs may be supposed weakly and meta-justified, as above?

It may be replied that the "internal" here is not internal in the right sense. What is internal in the right sense must remain restricted to the subjectivity of the subject, to that which pertains to the subject's psychology; it must not go outside of that, even to the physiological conditions holding in the subject's body; or at least it must not do so under the aspect of the physiological, even if in the end it is the physiological (or something physical anyhow) that "realizes" everything mental and psychological.

Even if we accept that objection, however, a very similar difficulty yet

remains for the conception of the blameless as the weakly justified or meta-justified (in either the weaker or the stronger sense). For it may be that the connections among the experiences and beliefs of the victim are purely random, as in the example above. True, in that example the randomness derives from the randomizing behavior of the demons involved. But there is no reason why the randomizing may not be brought inside. Thus, given a set of experiences or beliefs, there may be many alternative further beliefs that might be added by the subject, and there may be no rational mechanism that selects only one to be added. It may be rather that one of the many alternatives pops in at random: thus it is a radically random matter which alternative further belief is added in any specific case. Our evil demon's victim, though damaged internally in that way, so that his inner mental processes are largely random, may still by amazing coincidence acquire a coherent and comprehensive system of beliefs that makes him weakly justified and even meta-justified, in both the weaker and stronger senses indicated above. Yet is there not something still defective in such a victim, something that would preclude our holding him to be indiscernible from us in all internal respects of epistemic relevance?

Consider again the project of defining a notion of weak justification, however, a notion applicable to evil-demon victims in accordance with our intuitions; or that of defining a notion of meta-justification as above, one applicable equally to the victims and to ourselves in our normal beliefs. These projects may well be thought safe from the fact that a victim might be internally defective in ways that go beyond any matter of weak or meta-justification. Fair enough. But then of course we might have introduced a notion of superweak justification, and provided sufficient conditions for it as follows:

S is superweakly justified in a certain belief if (1) the cognitive process that produces the belief is unreliable, but (2) S has not acquired that belief as a result of a deliberate policy of acquiring false beliefs (a policy adopted perhaps at the behest of a cruel master, or out of a deep need for epistemic self-abasement).

Someone may propose that a similarity between the victim of the evil demon on one side and ourselves on the other is that we all are superweakly justified in our object-level beliefs in fires and the like. And this is fair and true enough. But it just does not go very far, not far enough. There is much else that is epistemically significant to the comparison between the victim and ourselves, much else that is left out of account by the mere notion of superweak justification. Perhaps part of what is left out is what the notion of weak justification would enable us to capture,

#### THEORIES OF JUSTIFICATION

and perhaps the notion of meta-justification, especially its stronger variant, would enable us to do even better. Even these stronger notions fall short of what is needed for fuller illumination, however, as I have tried to show above through the victims of randomization, whether demon-derived or internally derived. In order to deal with the new evil-demon problem and with the problem of meta-incoherence we need a stronger notion than either that of the weakly justified or that of the meta-justified, a stronger notion of the internally or subjectively justified.

#### D. A STRONGER NOTION OF THE “INTERNALLY JUSTIFIED”: INTELLECTUAL VIRTUE

Let us define an intellectual virtue or faculty as a competence in virtue of which one would mostly attain the truth and avoid error in a certain field of propositions F, when in certain conditions C. Subject S believes proposition P at time t out of intellectual virtue only if there is a field of propositions F, and there are conditions C, such that: (a) P is in F; (b) S is in C with respect to P; and (c) S would most likely be right if S believed a proposition X in field F when in conditions C with respect to X. Unlike Historical Reliabilism, this view does not require that there be a cognitive process leading to a belief in order for that belief to enjoy the strong justification required for constituting knowledge. Which is all to the good, since requiring such a process makes it hard to explain the justification for that paradigm of knowledge, the Cartesian cogito. There is a truth-conducive “faculty” through which everyone grasps their own existence at the moment of grasping. Indeed, what Descartes noticed about this faculty is its infallible reliability. But this requires that the existence which is grasped at a time t be existence at that very moment t. Grasp of earlier existence, no matter how near to the present, requires not the infallible cogito faculty, but a fallible faculty of memory. If we are to grant the cogito its due measure of justification, and to explain its exceptional epistemic status, we must allow faculties which operate instantaneously in the sense that the outcome belief is about the very moment of believing, and the conditions C are conditions about what obtains at that very moment – where we need place no necessary and general requirements about what went before.

By contrast with Historical Reliabilism, let us now work with intellectual virtues or faculties, defining their presence in a subject S by requiring

that, concerning propositions X in field F, once S were in conditions C with respect to X, S would most likely attain the truth and avoid error.



#### RELIABILISM AND INTELLECTUAL VIRTUE

In fact a faculty or virtue would normally be a fairly stable disposition on the part of a subject *relative to an environment*. Being in conditions C with respect to proposition X would range from just being conscious and entertaining X – as in the case of “I think” or “I am” – to seeing an object O in good light at a favorable angle and distance, and without obstruction, etc. – as in “This before me is white and round.” There is no restriction here to processes or to the internal. The conditions C and the field F may have much to do with the environment external to the subject: thus a moment ago we spoke of a C that involved seeing an external object in good light at a certain distance, etc. – all of which involves factors external to the subject.

Normally, we could hope to attain a conception of C and F which at best and at its most explicit will still have to rely heavily on the assumed nature of the subject and the assumed character of the environment. Thus it may appear to you that there is a round and white object before you and you may have reason to think that in conditions C (i.e., for middle-sized objects in daylight, at arm’s length) you would likely be right concerning propositions in field F (about their shapes and colors). But of course there are underlying reasons why you would most likely be right about such questions concerning such objects so placed. And these underlying reasons have to do with yourself and your intrinsic properties, largely your eyes and brain and nervous system; and they have to do also with the medium and the environment more generally, and its contents and properties at the time. A fuller, more explicit account of what is involved in having an intellectual virtue or faculty is therefore this:

Because subject S has a certain inner nature (I) and is placed in a certain environment (E), S would most likely be right on any proposition X in field F relative to which S stood in conditions C. S might be a human; I might involve possession of good eyes and a good nervous system including a brain in good order; E might include the surface of the earth with its relevant properties, within the parameters of variation experienced by humans over the centuries, or anyhow by subject S within his or her lifetime or within a certain more recent stretch of it; F might be a field of propositions specifying the colors or shapes of an object before S up to a certain level of determination and complexity (say greenness and squareness, but not chartreuseness or chiliagonicity); and C might be the conditions of S’s seeing such an object in good light at arm’s length and without obstructions.

If S believes a proposition X in field F, about the shape of a facing surface before him, and X is false, things might have gone wrong at

#### THEORIES OF JUSTIFICATION

interestingly different points. Thus the medium might have gone wrong unknown to the subject, and perhaps even unknowably to the subject; or something within the subject might have changed significantly: thus the lenses in the eyes of the subject might have become distorted, or the optic nerve might have become defective in ways important to shape recognition. If what goes wrong lies in the environment, that might prevent the subject from knowing what he believes, even if his belief were true, but there is a sense in which the subject would remain subjectively justified or anyhow virtuous in so believing. It is this sense of internal virtue that seems most significant for dealing with the new evil-demon argument and with the meta-incoherence objection. Weak justification and meta-justification are just two factors that bear on internal value, but there are others surely, as the earlier examples were designed to show – examples in which the experience/belief relation goes awry, or in which a randomizer gate intervenes. Can something more positive be said in explication of such internal intellectual virtue?

Intellectual virtue is something that resides in a subject, something relative to an environment – though in the limiting case, the environment may be null, as perhaps when one engages in armchair reflection and thus comes to justified belief.

A subject *S*'s intellectual virtue *V* relative to an "environment" *E* may be defined as *S*'s disposition to believe correctly propositions in a field *F* relative to which *S* stands in conditions *C*, in "environment" *E*.

It bears emphasis first of all that to be in a certain "environment" is *not* just a matter of having a certain spatio-temporal location, but is more a matter of having a complex set of properties, only some of which will be spatial or temporal. Secondly, we are interested of course in non-vacuous virtues, virtues which are not possessed simply because the subject would never be in conditions *C* relative to the propositions in *F*, or the like, though there may be no harm in allowing vacuous virtues to stand as trivial, uninteresting special cases.

Notice now that, so defined, for *S* to have a virtue *V* relative to an environment *E* at a time *t*, *S* does not have to be *in* *E* at *t* (i.e., *S* does not need to have the properties required). Further, suppose that, while outside environment *E* and while not in conditions *C* with respect to a proposition *X* in *F*, *S* still retains the virtue involved, *relative to E*, because the following ECF conditional remains true of *S*:

(ECF) that if in *E* and in *C* relative to *X* in *F*, then *S* would most likely be right in his belief or disbelief of *X*.

#### RELIABILISM AND INTELLECTUAL VIRTUE

If S does so retain that virtue in that way, it can only be due to some components or aspects of S's intrinsic nature I, for it is S's possessing I together with being in E and in C with respect to X in F that fully explains and gives rise to the relevant disposition on the part of S, namely the disposition to believe correctly and avoid error regarding X in F, when so characterized and circumstanced.

We may now distinguish between (a) possession of the virtue (relative to E) in the sense of possession of the disposition, i.e., in the sense that the appropriate complex and general conditional (ECF) indicated above is true of the subject with the virtue, and (b) possession of a certain ground or basis of the virtue, in the sense of possessing an inner nature I from which the truth of the ECF conditional derives in turn. Of course one and the same virtue might have several different alternative possible grounds or bases. Thus the disposition to roll down an incline if free at its top with a certain orientation, in a certain environment (gravity, etc.), may be grounded in the sphericity and rigidity of an object, or alternatively it may be grounded in its cylindricality and rigidity. Either way, the conditional will obtain and the object will have the relevant disposition to roll. Similarly, Earthians and Martians may both be endowed with sight, in the sense of having the ability to tell colors and shapes, etc., though the principles of the operation of Earthian sight may differ widely from the principles that apply to Martians, which would or might presumably derive from a difference in the inner structure of the two species of being.

What now makes a disposition (and the underlying inner structure or nature that grounds it) an intellectual virtue? If we view such a disposition as defined by a C-F pair, then a being might have the disposition to be right with respect to propositions in field F when in conditions C with respect to them, relative to one environment E but not relative to another environment E'. Such virtues, then, i.e., such C-F dispositions, might be virtuous only relative to an environment E and not relative to a different environment E'. And what makes such a disposition a virtue relative to an environment E seems now as obvious as it is that having the truth is an epistemic desideratum, and that being so constituted that one would most likely attain the truth in a certain field in a certain environment, when in certain conditions *vis-à-vis* propositions in that field, is so far as it goes an epistemic desideratum, an intellectual virtue.

What makes a subject intellectually virtuous? What makes her inner nature meritorious? Surely we can't require that a being have all merit and virtue before it can have any. Consider then a subject who has a minimal virtue of responding, thermometer-like, to environing food, and suppose him to have the minimal complexity and sophistication required for having beliefs at all – so that he is not literally just a thermometer or the

like. Yet we suppose him further to have no way of relating what he senses, and his sensing of it, to a wider view of things that will explain it all, that will enable him perhaps to make related predictions and exercise related control. No, this ability is a relatively isolated phenomenon to which the subject yields with infant-like, unselfconscious simplicity. Suppose indeed the subject is just an infant or a higher animal. Can we allow that he knows of the presence of food when he has a correct belief to that effect? Well, the subject may of course have reliable belief that there is something edible there, without having a belief as reliable as that of a normal, well-informed adult, with some knowledge of food composition, basic nutrition, basic perception, etc., and who can at least implicitly interrelate these matters for a relatively much more coherent and complete view of the matter and related matters. Edibility can be a fairly complex matter, and how we have perceptual access to that property can also be rather involved, and the more one knows about the various factors whose interrelation yields the perceptible edibility of something before one, presumably the more reliable one's access to that all-important property.

Here then is one proposal on what makes one's belief that-p a result of enough virtue to make one internally justified in that belief. First of all we need to relativize to an assumed environment, which need not be the environment that the believer actually is in. What is required for a subject S to believe that-p out of sufficient virtue relative to environment E is that the proposition that-p be in a field F and that S be in conditions C with respect to that proposition, such that S would not be in C with respect to a proposition in F while in environment E, without S being most likely to believe correctly with regard to that proposition; and further that by comparison with epistemic group G, S is not grossly defective in ability to detect thus the truth in field F; i.e., it cannot be that S would have, by comparison with G:

- (a) only a relatively very low probability of success,
- (b) in a relatively very restricted class F,
- (c) in a very restricted environment E,
- (d) in conditions C that are relatively infrequent,

where all this relativity holds with respect to fellow members of G and to their normal environment and circumstances. (There is of course some variation from context to context as to what the relevant group might be when one engages in discussion of whether or not some subject knows something or is at least justified in believing it. But normally a certain group will stand out, with humanity being the default value.)

## E. INTELLECTUAL VIRTUE APPLIED

Consider now again the new evil-demon problem and the problem of meta-incoherence. The crucial question in each case seems to be that of the internal justification of the subject, and this in turn seems not a matter of his superweak or weak or meta justification, so much as a matter of the virtue and total internal justification of that subject relative to an assumed group G and environment E, which absent any sign to the contrary one would take to be the group of humans in a normal human environment for the sort of question under consideration. Given these assumptions, the victim of the evil demon is virtuous and internally justified in every relevant respect, and not just in the respects of enjoying superweak, weak, and meta justification; for the victim is supposed to be just like an arbitrarily selected normal human in all cognitively relevant internal respects. Therefore, the internal structure and goings on in the victim must be at least up to par, in respect of how virtuous all of that internal nature makes the victim, relative to a normal one of us in our usual environment for considering whether we have a fire before us or the like. For those inclined towards mentalism or towards some broadly Cartesian view of the self and her mental life, this means at a minimum that the experience-belief mechanisms must not be random, but must rather be systematically truth-conducive, and that the subject must attain some minimum of coherent perspective on her own situation in the relevant environment, and on her modes of reliable access to information about that environment. Consider next those inclined towards naturalism, who hold the person to be either just a physical organism, or some physical part of an organism, or to be anyhow constituted essentially by some such physical entity; for these it would be required that the relevant physical being identical with or constitutive of the subject, in the situation in question, must not be defective in cognitively relevant internal respects; which would mean, among other things, that the subject would acquire beliefs about the colors or shapes of facing surfaces only under appropriate prompting at the relevant surfaces of the relevant visual organs (and not, e.g., through direct manipulation of the brain by some internal randomizing device).<sup>12</sup>

We have appealed to an intuitive distinction between what is intrinsic or internal to a subject or being, and what is extrinsic or external. Now when a subject receives certain inputs and emits as output a certain belief or a certain choice, that belief or choice can be defective either in virtue of

<sup>12</sup> As for the generality problem, my own proposed solution appears in Chapter 16 of this volume.

## THEORIES OF JUSTIFICATION

an internal factor or in virtue of an external factor (or, of course, both). That is to say, it may be that everything inner, intrinsic, or internal to the subject operates flawlessly and indeed brilliantly, but that something goes awry – with the belief, which turns out to be false, or with the choice, which turns out to be disastrous – because of some factor that, with respect to that subject, is outer, extrinsic, or external.<sup>13</sup>

In terms of that distinction, the victim of the demon may be seen to be internally justified, just as internally justified as we are, whereas the meta-incoherent are internally unjustified, unlike us.

My proposal is that justification is relative to environment. Relative to our actual environment A, our automatic experience-belief mechanisms count as virtues that yield much truth and justification. Of course relative to the demonic environment D such mechanisms are not virtuous and yield neither truth nor justification. It follows that relative to D the demon's victims are not justified, and yet *relative to A their beliefs are justified*. Thus may we fit our surface intuitions about such victims: that they lack knowledge but not justification.

In fact, a fuller account should distinguish between “justification” and “aptness”<sup>14</sup> as follows:

(a) The “justification” of a belief B requires that B have a basis in its inference or coherence relations to other beliefs in the believer's mind – as in the “justification” of a belief derived from deeper principles, and thus “justified,” or the “justification” of a belief adopted through cognizance of its according with the subject's principles, including principles as to what beliefs are permissible in the circumstances as viewed by that subject.

(b) The “aptness” of a belief B relative to an environment E requires that B derive from what relative to E is an intellectual virtue, i.e., a way of arriving at belief that yields an appropriate preponderance of truth over error (in the field of propositions in question, in the sort of context involved).

As far as I can see, however, the basic points would remain within the more complex picture as well. And note that “justification” itself would

13 This sort of distinction between the internal virtue of a subject and his or her (favorable or unfavorable) circumstances is drawn in “How Do You Know?” – Chapter 2 in this volume. There knowledge is relativized to epistemic community, though not in a way that imports any subjectivism or conventionalism, and consequences are drawn for the circumstances within which praise or blame is appropriate (see especially the first part of Section II).

14 For this sort of distinction, see, e.g., “Methodology and Apt Belief,” Chapter 14 in this volume. The more generic distinction between external and internal justification may be found in “The Analysis of ‘Knowledge That P’,” Chapter 1 in this volume.

then amount to a sort of inner coherence, something that the demon's victims can obviously have despite their cognitively hostile environment, but also something that will earn them praise relative to that environment only if it is not just an inner drive for greater and greater explanatory comprehensiveness, a drive which leads nowhere but to a more and more complex tissue of falsehoods. If we believe our world not to be such a world, then we can say that, relative to our actual environment A, "justification" as inner coherence earns its honorific status, and is an intellectual virtue, dear to the scientist, the philosopher, and the detective. Relative to the demon's D, therefore, the victim's belief may be inapt and even unjustified – if "justification" is essentially honorific – or if "justified" simply because coherent then, relative to D, that justification may yet have little or no cognitive worth. Even so, relative to our environment A, the beliefs of the demon's victim may still be both apt and valuably justified through their inner coherence.

The epistemology defended in this volume – virtue perspectivism – is distinguished from generic reliabilism in three main respects:

- (a) Virtue perspectivism requires not just any reliable mechanism of belief acquisition for belief that can qualify as knowledge; it requires the belief to derive from an intellectual virtue or faculty (a notion defined more fully in Chapter 16).
- (b) Virtue perspectivism distinguishes between aptness and justification of belief, where a belief is apt if it derives from a faculty or virtue, but is justified only if it fits coherently within the epistemic perspective of the believer – perhaps by being connected to adequate reasons in the mind of the believer in such a way that the believer follows adequate or even impeccable intellectual procedure (see Chapter 14). This distinction is used as one way to deal with the new evil-demon problem. (See Chapter 16, Section D.)
- (c) Virtue perspectivism distinguishes between animal and reflective knowledge. For animal knowledge one needs only belief that is apt and derives from an intellectual virtue or faculty. By contrast, reflective knowledge always requires belief that not only is apt but also has a kind of justification, since it must be belief that fits coherently within the epistemic perspective of the believer (see Chapter 13, Section IX). This distinction is used earlier in this chapter to deal with the metaincoherence problem, and it also opens the way to a solution for the generality problem. (See Chapter 16, Section D.)