This is a late draft of a paper that appeared in the

Croatian Journal of Philosophy (2006), 190-221.

Transglobal Reliabilism

David Henderson and Terry Horgan

University of Memphis and University of Arizona

1. Introduction.

We here propose an account of what it is for an agent to be objectively justified in holding some belief. We present in outline this approach, which we call *transglobal reliabilism,* and we discuss how it is motivated by various thought experiments. While transglobal reliabililsm is an externalist epistemology, we think that is accommodates traditional internalist concerns and objections is a uniquely natural and respectful way.[1]

Transglobal reliabilism is a species of reliabilism. It holds that it is necessary for being objectively justified that the belief in question be fixed by way of processes that are reliable—where reliability, as always, must be understood as relative to some reference class of environments. Transglobal reliabilism is distinguished by its conception of the relevant reference class of environments. The more familiar versions of reliabilism treat the relevant form of reliability as reliability relative the agent's own actual global environment. For transglobal reliabilism, the relevant form of reliability is reliability relative to the set of *experientially possible global environments—transglobal reliability*. A possible global environment is *experientially possible* just in case it is compatible with one's there having experiences of roughly the character of those that agents actually have.[2] The actual global environment is but one among a diversity of experientially possible global environments. Some experientially possible global environments would be extremely epistemically inhospitable—there would be few if any globally reliable processes to be had there. Demon-infested global environments, and those featuring agents as envatted brains, would be cases in point. The actual global environment (which we take to be demon free) is a moderately hospitable experientially possible global environment; there would seem to be both more and less hospitable global environments. In yet more hospitable experientially possible global environments, there would be fewer sources of error than there are in the actual global environment. Transglobal reliabilism, in a somewhat simplified nutshell, is the view that it is constitutively required for objective epistemic justification that a belief be fixed by processes that are transglobally reliable—reliable relative to the class of experientially possible global environments.

It should be noted that for a process to be transglobally reliable does not require that the process be globally reliable in all the experientially possible global environments. Rather, and again somewhat crudely, it requires that the belief-fixing process be reliable in a wide range of such global environments. Consider the parallel matter involving what might be termed *local* versus global reliability. Local reliability is reliability relative to some local environment that an agent encounters within that agent's actual global environment. Within the global environment, there are relatively epistemically hostile and relatively hospitable local environments. Were there a fake-barn county within the agent's global environment, it would constitute a relatively inhospitable local environment vis-à-vis the formation of perceptually based beliefs about barnhood. Obviously, within our actual global environment, there are local environments that are epistemically relatively inhospitable (thus, the fog of war, the fog of American politics, the fog of fundamentalism, the fog of hypoxia at altitude in a white-out, and so on). There are also relatively hospitable local environments (presumably one's own kitchen is relatively hospitable for perception of everyday objects, with its lack of camouflage, with its paucity of fakes, and with the prevalence of familiar objects, and oxygen). It is clear that a process can be globally reliable while failing to be locally reliable in some inhospitable local environments. Similarly, a process can be transglobally reliable while failing to be globally reliable in all experientially possible global environments.

In what follows we will consider a range of thought experiments. We suggest that these reveal a deep and important epistemic concern: that one's beliefs be produced and maintained by processes that are *epistemically safe*. That is, judgments about objective epistemic justification having to do with a range of scenarios reflects a concern for the epistemic safety of agent's processes. This strongly suggest that such safety is conceptually required for objective epistemic justification. In effect, an adequate account of objective epistemic justification must then provide an account of what makes for such safety. We argue that the understanding of epistemic safety that is afforded by transglobal reliabilism is a superior to that provided by classical reliabilism. It clarifies the demand that, in order to be justified, beliefs must be generated or maintained in ways that are *not unduly or needlessly risky*—a demand reflected in common judgment tendencies regarding the scenarios now to be considered.

2. A Series of Thought Experiments: The Road to Transglobal Reliabilism

The motivation for transglobal reliabilism to be presented here proceeds in two stages. First, the thought experiments considered in section 2.1 reveal a pivotal concern for epistemic safety—one served by employing processes with robust forms of reliability. Second, in section 2.2, as we consider further thought experiments, it will become clear that the indicated concern for epistemic safety can best be accommodated by an account of objective epistemic justification that gives a pivotal place to transglobal reliability. Thus, the thought experiments considered in 2.1, ultimately (when considered along with those to be considered in 2.2) provide motivation for transglobal reliabilism. First, however, we consider them as supporting a more familiar variant of reliabilism.

2.1. Classical and neoclassical (or global) reliabilism.

A generic form of reliabilism has become fairly standard. It emphasizes reliability within the worldly environment that the epistemic agent *actually occupies*—as opposed to various non-actual worldly environments such as that of an envatted brain. Adherents of this general position—call it *classical reliabilism*—have commonly not distinguished sharply, or not made much of, the distinction between local reliability and global reliability. In this section we consider scenarios suggesting that, were the choice between just these two forms of reliability, then reliabilist accounts of objective justification would best focus on global reliability. We call the refined reliabilism thus suggested *neoclassical reliabilism* or *global reliabilism.*

a. Athena and Fortuna

Suppose that Athena and Fortuna are driving from New York City to Memphis. In rural West Virginia, they drive through a county in which there happen to be numerous extremely realistic-looking fake barns within view—although neither of them has any inkling of this fact or any reason to suspect. As it happens, in this local area all the real barns are yellow, and none of the fake barns or any other buildings are yellow. Again, they have no information to this effect. As they drive past a saliently presented yellow building, Athena, who has had reasonable experience with barns, gets a clear look at it, and on the basis of its barn-like visual appearance, she judges it to be a barn.

Fortuna gets only a very brief glimpse of the building. She saw her first barn just yesterday, elsewhere, and it happened to be yellow. She judges, on the basis of the briefly-glimpsed building's yellow color, that it is a barn—even though she did not get a good look at it, and was thus unable to discern any features that are generally distinctive of barns as opposed to other kinds of buildings. It's not that she has a *general* belief that all and only yellow buildings are barns, or that all barns are yellow; she has never formed any such belief. Also, it's not that she has a *general* tendency to inductively extrapolate from old cases to new ones in a hastily-generalizing way, a tendency she might otherwise be exhibiting here. Rather, it just happens that *in the present circumstances*, a psychological process is present within her that takes as input both the brief glimpse of a yellow building and the yellow-barn memory, and generates as output the barn-belief about the briefly glimpsed object.

So Athena and Fortuna each form the belief that the building is a barn. And indeed it is.

First consider Athena. One's strong inclination is to say that her belief that the building is a barn is extremely well justified. After all, she has excellent perceptual evidence for the belief, and she formed the belief using a process of perceptual barn-categorization that generally works extremely well. On the other hand, as is commonly stressed in discussions of fake-barn scenarios, one is also strongly inclined to say that she does not *know* that the building is a barn, because her belief has been produced by a belief-forming process that happens to be *unreliable in this specific local environment*. Given all those fake barns around, any of which she would have mistakenly taken to be a real barn, the truth of her belief is too much a matter of epistemic luck to count as knowledge.

Now consider Fortuna. Here one has a very strong inclination to say that she *lacks* objective justification for her belief that the building glimpsed is a barn. This seems so despite the fact that this belief was produced by a belief-forming process that happens to be reliable in the local vicinity she then occupies. The trouble is that *this reliability is itself too fortuitous*, too much a matter of epistemic luck, for the belief to count as justified. And as a consequence, the truth of her belief is also too much a matter of luck for the belief to count as *knowledge*—the reliability of the belief-forming process notwithstanding.

b. Local Reliability, Global Reliability, and the Fake-Barn Scenario

If one takes these intuitive verdicts to be correct, would this mean that we have here a case in which one person's belief (Athena's) is justified even though it is *not* produced by a reliable belief-forming process, and in which another person's belief (Fortuna's) fails to be justified even though it *is* produced by a reliable belief-forming process? That would be too flat-footed a moral to draw; one cannot flatly say that the relevant process is reliable, or that it is not reliable.

Athena employs a barn-categorizing belief-forming process that is indeed highly reliable in one respect: it is globally reliable. Recall that she has had reasonable experience with barns, and it is plausible to suppose that she has thereby come to have a reliable perceptual ability to discriminate barns from non-barns. (In the world at large, there are few fake barns. Humans with reasonable experience typically do become reliable perceptual judges with respect to such generally non-tricky perceptual matters involving common middle-sized physical objects.) But her barn-categorizing process is locally unreliable, because of all those fake barns in the vicinity. Intuitively, the strong objective justification her belief possesses does indeed involve production by a reliable belief-forming process—viz., a globally reliable one. On the other hand, intuitively, the fact that this strongly justified belief nonetheless fails to qualify as knowledge also involves a failure with respect to reliability of the belief-forming process—viz., a failure to be locally reliable. Both dimensions of reliability thus figure importantly: the presence of global reliability seems to figures in her belief being well justified, whereas the absence of local reliability seems to figure in that belief nevertheless failing to qualify as a case of knowledge. At least these would seem to provide a reasonable suggestion for making sense of our judgments here.

Fortuna is a converse case. Her belief is produced by a barn-categorization process that is locally reliable (for reasons beyond her cognitive ken) but is globally unreliable. One judges that

Fortuna's belief fails to be well justified. Intuitively, what appears to figure importantly in its lacking objective justification is the fact that the belief-forming process is not globally reliable. Given this lack of global reliability, the fact that the process happens to be locally reliable strikes one intuitively as a *lucky accident*, epistemically speaking; and for this reason, the belief also does not count as a case of knowledge.

c. Global Reliability as a Robust Disposition

Call a disposition *robust* if it obtains relative to a fairly wide reference class of potential circumstances, situations, or environments. A robust disposition is one whose possession does not depend heavily upon certain unusual or atypical features that are highly specific to the particular circumstance or environment which the possessor of the disposition might happen to occupy; i.e., the disposition does not obtain only relative to a narrow reference class of environments in which those particular features happen to be present.

Now consider reliability, with robustness in mind. The tendency to produce (mostly) true beliefs must be understood as relative to a reference class of actual or possible environments. If a process is globally reliable, it has this tendency with respect to the wide reference class comprising the potential local environments to which an agent might be exposed (or which the agent might inhabit) within that agent's global environment. This is to have reliability in a reasonably robust fashion. For a process to be globally reliable is for its reliability not to depend heavily upon certain unusual or atypical features that are highly specific to the particular circumstance or environment which the possessor of the disposition might happen to occupy; its reliability then does not obtain only relative to a narrow reference class of environments in which those particular features happen to be present. This said, a process can be globally reliable while failing to be locally reliable with respect to some local environment afforded by the global environment (as illustrated by Athena's processes), and a process can be locally reliable without being globally reliable (as illustrated by Fortuna's processes). When a process is locally, but not globally reliable, that process's reliability does depend heavily upon certain unusual or atypical features that are highly specific to the particular circumstance or environment which the possessor of the disposition occupies. Such merely local reliability is non-robust, because it involves a narrow reference class: local environments with highly specific features exhibited by the particular local environment that an agent happens to be in.

Compare automobiles. An old car with its somewhat compromised cooling system, worn tires, and the like, might well be reliable relative to a narrow range of potential environmental circumstances, but not relative to a wide range. It might be reliable in a local environment where the climate is temperate, it rains moderately and seldom snows, there are no long steep hills, and there is very little traffic. But it might be unreliable in local environments where there are temperature extremes, or lots of traffic, or demanding hills, or significant of rain, snow and ice, etc. Although the car might be in service in an environment in which it happens to be locally reliable, it is not globally reliable. It thus is not *robustly* reliable. Its reliability depends heavily on the somewhat unusual combination of features particular to the environment where it happens to be employed. On the other hand, there are fine new cars that qualify as globally reliable—that is, would stand up well to the demands of providing daily transportation with respect to the representative environments to which an auto might be exposed or put to work. Of course, such cars would fail in a few local environments—for example, those characterized by significant flooding, high electromagnetic surges (associated with nuclear explosions) or volcanic flows. But, with respect to the global environment taken as a whole, as a composite of the range of potential local environments presented within the global environment, the relevant cars would tend to provide daily transportation on demand. Although a robustly reliable item—a car, a belief-forming process, or whatever—is reliable in a *wide range* of the potential circumstances within the reference class of the item's reliability, there can be *exceptional* local circumstances within which the item fails to be locally reliable. This is a ubiquitous feature of robust dispositions.

Here is a useful way of thinking about or intuitively gauging the robustness of the reliability of a process or item, a way of doing so that will prove useful as we proceed. Start with a process or item that is reliable in a given environment (for now, think of a particular local environment). Then think of varying the environment in various ways. To continue the automotive example, compare two vehicles: the old car mentioned above and a new car of a make that would be highly rated in measures like the *Consumer Reports* survey of automotive reliability. (Because respondents to that survey are widely distributed geographically, it may be taken to measure global reliability.) Both vehicles are locally reliable in the benign environment described above. So, begin varying that environment. As one considers variations in temperatures typically encountered in the environment, one finds that the older vehicle quickly ceases to the reliable, while the newer vehicle remains reliable though more variation. Similar results are obtained when considering variations in precipitation type and amount, or variations in the length or degree of inclines to be managed. The vehicle that retains its reliability across greater variations in environmental conditions is more robustly reliable. As we noted, robustness of reliability is a matter of not being highly dependent on particular environmental conditions—a matter of accommodating more rather than less environmental variation.

d. Epistemic Safety: a Constitutive Requirement for Justification

Return again to Fortuna, and consider what seems intuitively objectionable about her belief that the yellow structure she fleetingly glimpsed is a barn. The problem is that, although the process that generated this belief does happen to be locally reliable, its being so is too much a matter of epistemic *luck* for her belief to be justified. Adopting the belief in the way she did—on the basis of a brief and fleeting glance in which the only salient feature she noticed was the structure's color, and in a way that arises somehow out of the fact that the one barn she has previously seen was yellow—is objectionably risky, epistemologically speaking. It is not epistemically *safe*. Somehow Fortuna has deployed a process that is not at all globally reliable, and rather accidentally or coincidentally, she just happened to do so in a rather special local environment in which that process was locally reliable. This seems an extraordinary lucky coincidence, for there is nothing about Fortuna that in any way "tracks" the features of the local environment in virtue of which her yellow-building triggered barn-identification generating process is locally reliable. It is precisely because the manner of belief-formation is so unsafe that its local reliability is too much a matter of luck to count for much, insofar as justification is concerned.

Compare the epistemically more delightful Athena. She, like we in our everyday lives, employs the sort of perceptual process that has been shaped by reasonable courses of experience with common enough objects. Experientially informed, or "trained up," perceptual processes having to do with familar objects in clear view qualify as globally reliable processes. It strikes one as relatively safe to employ such a process in the absence of information suggesting that conditions are somehow exceptional.

Epistemic safety, then, is clearly associated with being a *suitably* truth-conducive belief-forming process. At least in the case of Athena and Fortuna, it is *global reliability that makes for the difference that one senses in epistemic safety*. One might also put the contrast, and the difference in epistemic safety, in terms of the differential robustness of the reliability of the processes in play.

One can begin to appreciate why the epistemic safety—and the relative robustness of reliability—of belief-fixing processes would be epistemologically significant when one reflects on a prominent and pervasive characteristic of epistemic life: one's epistemic endeavor *must* be undertaken in the face of uncertainty. One's epistemic chores must be managed while possessing only a fallible understanding regarding one's global and local environment. Such epistemic uncertainty (or fallibility) regarding one's environment is paralleled by, or mirrored in, an uncertainty (or fallibility) regarding what processes will work in one's environment. In view of the uncertainty characteristic of the epistemic situation, consider two alternatives illustrated in the above discussion. On the one hand, one might employ a process that is reliable with respect to the wide reference class comprising the potential local environments to which one might be exposed (or which one might inhabit) within one's global environment. In other words, one might employ a globally reliable process; such a processes is reasonably robustly reliable, as explained already. It is relatively safe insofar as its reliability then does not depend heavily upon certain unusual or atypical features that are highly specific to the particular circumstance or environment which the possessor of the disposition might happen to occupy; its reliability then does not obtain only relative to a narrow reference class of environments in which those particular features happen to be present. Alternatively, one might employ processes that would be reliable only given certain unusual or atypical environmental features—only given features that are highly specific to particular circumstances or environments within the global environment that one happens to occupy.[3] *If*, in employing such a process one employs a reliable process, it is a merely locally reliable process. In light of the uncertainty that is a pervasive fact of epistemic life, it is clear what one should make of this general abstract choice: the safety afforded by globally reliable processes would be rationally desirable in preference to the risk one runs when using a merely locally reliable process. Globally reliable processes are safer and epistemically more valuable insofar as they afford one a certain margin of error (or margin for ignorance) regarding what environment within one's actual global environment one happens to occupy. Because the reliability of a globally reliable process is relatively robust, not so highly dependent on unusual features of one's local environment, it is relatively safe.

A natural line of thought, at this point, represents what we believe is a significant insight into what is constitutively required for objective epistemic justification. The thought we now suggest is fully consonant with the above reflections on Athena and Fortuna, local and global reliability, and epistemic safety and risk—further, it is a line of thought that we think is well within the spirit of standard reliabilism:

> First, embrace the idea that safety—a feature as yet to be more fully explained—is indeed a *constitutive requirement* a belief's being objectively justified.
>
> Second, look to give an account of safety as the belief's having been produced by some *suitably robustly reliable* belief-forming process.

We suggest that this line of thought provides the seeds of a very workable understanding of objective epistemic justification.

e. The Need for Further Refinement: the Case of Diana and Elena

How should one understand this idea of a suitably robustly reliable process? One might think, with an eye on the Athena/Fortuna cases, that global reliability *per se* is just the ticket. But a moment's reflection reveals that this proposal is too crude as it stands. Consider, for instance, someone—we will name her Diana—who has knowledge that the particular local region in question is full of very realistic fake barns, and of the location of this region. She has read about it in the local morning paper, having spent the night a hotel along the interstate located just a few miles from the region in question. She now finds herself driving through that peculiar local region. It remains true that her perceptual barn-categorization process is *globally* reliable; after all, her global environment is the same as Athena's, and is as it was before she encountered the local tourist information, so those processes remain globally reliable. Still, Diana most certainly would *not* be justified in coming to believe of a barn-looking structure off in a field, on the basis of its visual appearance, that it's a barn. After all, she believes with excellent reason that, in her current local environment, the reliability of this belief-forming process is compromised (it is locally unreliable). Considered of themselves, those perceptual processes remain globally reliable, here are locally unreliable, but in this case would not give rise to objectively justified beliefs (unlike in the case of Athena).

Likewise, consider the case of Elena. Suppose she knows, concerning that particular local area, both (a) that there are lots of barn facsimiles, and also (b) that all genuine barns, and no other structures of any kind (including both other kinds of buildings and barn facsimiles) are bright yellow. Finding herself in that very location (and knowing it), she catches a fleeting glance of a bright yellow structure, and promptly forms the belief that it's a barn. Now, of itself, such a process of barn-categorization, on the basis of glimpsed yellow color, would be globally *un*-reliable. But, now, unlike the case of Fortuna, although Elena formed her belief on the basis of a globally unreliable process (viz., classifying briefly glimpsed structures as a barns on the basis of their bright yellow color), one judges that this belief of hers *is* justified, even so. Such processes here are locally reliable, globally unreliable, and yet here feature in the production of objectively justified beliefs (unlike in the case of Fortuna). What is going on?

So there's a task to be undertaken in forging a neoclassical version of reliabilism that will incorporate the idea of safety of belief-forming processes as a constitutive requirement of justification: one needs to give an account of such safety in terms of some suitably robust form of reliability. The cases of Athena and Fortuna reflected the importance of safety, and provided some reason to think of it in terms of global reliability. But, as just demonstrated by the cases of Diana and Elena, global reliability *per se* won't fill the bill. But perhaps some refinement of it will. Notably, what seems different in the two sets of cases is how the processes in play—the perceptual barn-categorizing processes—themselves are related to further information and processes. So the suggestion to be developed is that epistemic safety might be understood in terms of suitably robust reliability, and that this will at least sometimes turn on how processes can be conditioned or modulated by certain other processes and the information thereby provided.

f. A Refinement: Suitable Modulational Control

An important general point emerges when reflecting on the cases of Diana and Elena: Cognitive agents like humans deploy various belief-forming processes in ways that are *holistically integrated* within the agent's overall cognitive architecture. Very frequently, such processes are employed not in isolation, but rather *under the control of various other or wider cognitive processes that condition or modulate them*. In such cases, the application or implementation of the one process is modulated in ways that are informed by the agent's wider cognitive processes. When these modulating or controlling cognitive processes provide veridical information about the agent's local environments, such control or modulation makes for a selective application of the modulated process, or otherwise tailors its application to aspects of those local environments about which information is had—thereby enhancing the reliability of the process so conditioned. We can then say that the process *is attuned* (the success term, 'attuned', and variations, 'attunement' and 'being attuned', are here used to indicate an epistemically good result). In principle, a whole host of different conditioning or modulating relations might be epistemically important. The wider processes might give rise to a narrower process—designing it or otherwise selecting or spawning it. They might selectively trigger the conditioned process in ways that are fitting, or thought to be appropriate. They might inhibit it—making for a more selective use of the process. They might spawn or otherwise cobble together a process to be triggered. All such modulation can and should be found among normal human cognitive agents.

Attunement, the actual modulation of one process by wider processes—where those wider processes have had the occasion to come by certain significant information bearing on the reliability of the modulated processes—can take time. It is too much to demand for objective justification that such attunement has been achieved. Athena did not come across the information that Diana encountered. Her processes were not attuned in the way Diana's came to be. But one still judges that she was justified in her belief—provided she *would have* refined her processes in the way Diana does, were she to have encountered that information. What then seems required for objective justification is thus more a matter of having appropriate control processes in place than a matter of those processes having had the occasion for effecting refinements or attunement.

A belief forming process P may be under the conditioning control of a wider set of processes—with or without those wider processes having yet come by information that prompts changes in, or modulations of, P. When there is such a functional-dependence relationship between processes, we will say that the process P is *under the modulatory control of* the wider processes. This wider set of processes may be termed *conditioning processes* with respect to P. So, belief-forming process P is *under the modulational control of* a wider set of processes S within the agent's cognitive system, provided that S would tailor P (would trigger P, inhibit P, or the like), were S to come to generate or possess certain relevant information.

g. The Suggested Position: Neoclassical Reliabilism

Conditioning processes themselves can be reliable or unreliable, can exhibit either or both of the two kinds reliability (local and global), and can be reliable (in either of these ways) to greater or lesser degrees. A fundamentally important feature that conditioning processes presumably ought to possess is that they be globally reliable *themselves*. There are various ways of thinking about this matter. For present purposes, it is sufficient to connect this desirability to the idea of epistemic safety. We have found that the robustness of reliability of a belief-fixing process enhances the epistemic safety of that process and the beliefs that it spawns—reduces the riskiness of one's beliefs, from an epistemic point of view. Belief-fixing processes that are merely locally reliable are unacceptably risky; those with global reliability are better off epistemologically. In a parallel fashion, if the conditioning processes are merely local reliable, this would seem to render the modulated processes themselves unacceptably risky. Typically, if the conditioning processes are merely locally reliable, then whatever benefits in terms of reliability, or robustness of reliability, is there provided by modulation by them, these gains are a matter of epistemic luck.[4] Modulation by globally reliable modulating processes is clearly less risky. Modulation by conditioning processes which are not themselves robustly reliable is something of a "house of cards"—risky, even when it just happens to "stand up" in a local context.

Putting together the points advanced in this section, one arrives at a refined form of classical reliabilism—viz., neoclassical reliabilism (or *global* reliabilism). The proposal is this:

for an agent to be justified in a belief, the belief must be formed or maintained by an epistemically safe cognitive process, where a belief-forming process P is safe just in case it is *globally reliable under suitable modulational control*.

2.2. Transglobal Reliabilism

We now discuss a few thought experiments indicating that global reliability under suitable modulational control does not yield an adequate account of objective epistemic justification. However, these thought experiments do not suggest that epistemic safety is not needed for objective justification, nor do they indicate that safety is not afforded by robust reliability. Rather, these lessons from the foregoing are preserved and extended. Instead, the thought experiments we will now discuss suggest that the needed safety is provided by a yet more robust form of reliability, different from global reliability.

a. Epistemic Safety without Suitably Modulated Global Reliability: The New Evil Demon Problem

Lehrer and Cohen (1985) formulated an objection to reliabilism that Sosa (1991a) has labeled the "new evil demon problem." In keeping with philosophical mythology, suppose that there is an evil demon—malicious and very powerful—out to deceive the agent. Seeking to epistemically defeat the agent at every turn, the demon provides the agent with appearances or experiences that seem to indicate a compellingly consistent environment—at least as compellingly consistent as the global environment that we (we and our readers) inhabit. As the agent undertakes to do things in that environment, the demon responds by giving the agent the fitting appearances. But the environment that the agent and the demon inhabit is radically other than the environment that the agent is led to imagine and theorize about—thus the deception. Further, suppose that the cagey demon stands ready to adapt to whatever processes the agent might employ—so that, given the processes employed, the agent will be given input that will (in combination with those processes) lead him or her wrong. The demon is here capable of defeating the agent no matter what processes the agent adopts. These days, computers are commonly cited in place of evil-demons when constructing skeptical scenarios. Typically, one supposes that the supercomputer has charge of a brain in a vat.

Whether plagued by evil demons or envatted, an agent in such an exceedingly inhospitable global environment is doomed to a kind of epistemological failure. There are no reliable processes to be had in such an environment. Whatever cognitive processes the agent employs, the demon will counter with input that is fittingly deceptive (that leads to false beliefs) given those processes.[5] The agent will end with systematically false beliefs.

Consider several agents inhabiting such a demon infested environment. To begin with, we may suppose that some agent, call her Constance, is remarkably like the intelligent, educated, and conscientious, epistemic agents one would want to include in one's own epistemic community. She avoids fallacious ways of reasoning, both deductive and inductive. Constance avoids inconsistency as well as the best of us. She maintains a high degree of wide reflective equilibrium in her belief system(s). In keeping with the facts as just stipulated, she only generalizes when samples are large enough for statistical confidence at some high level, and then only when the samples are either random or characterized by a diversity that seems to match the distribution of likely causal features in the population. Further, Constance has taken note of where her observations have seemed untrustworthy in the past, and discounts certain observations accordingly. In crucial respects, she is like the best perceptual agents among us. She is like those who, on the basis of long experiential refinement, prompted by successes and failures, have come to be sensitive, careful, and discriminating perceivers of everyday things. Constance has become trained up through a long process of perceptual refinement, drawing on apparent successes and failures, to be a sensitive and careful perceptual agent.[6] Put simply, if one were picking epistemic teams to play in our actual global environment, one would not hesitate to pick Constance for one's own team, as long as she could join us in our actual global environment. In the actual environment, Constance would be a model epistemic citizen.

Suppose then that Constance, in her demon infested environment, hears familiar noises clearly emanating from the phone on her desk (or so things appear), and that she then forms the belief that someone is calling. Of course, the belief is false and arises by way of a highly unreliable process. All processes are both locally and globally unreliable in an environment where a powerful and resourceful evil demon (or analogous supercomputer) is at work on the inhabitants. Still, there is a very strong tendency to judge that Constance is objectively justified in holding this belief. Of course, one does not find this scenario to be a happy or epistemically desirable scenario. But, one is strongly disinclined to find fault with the agent. The problem lies with the agent's extremely inhospitable environment—its demon-infestation. There, no agent and no process will help. So, one judges that the problem is in with the environment, and not the agent. Fine agent, lousy environment. In keeping with this, one judges that Constance is justified and has done nothing epistemically wrong or inappropriate in believing that someone is calling. [7]

This judgment, poses a clear challenge to standard reliabilist accounts of justification. If it is honored, then such accounts must be abandoned. Neoclassical reliabilism is not immune to the challenge. This refined position takes global reliability under suitable globally-reliable modulational control to be constituitively required for objective justification—to be necessary for justification. Notice that Constance has in place very significant conditioning processes, exhibiting significant modulational control (of a sort that in less inhospitable environments would be globally reliable themselves, and would suitably enhance the global reliability of the processes they condition), but these cannot contribute to global reliability in her global environment, for no processes will avail agents in this environment. Constance's processes lack the property that neoclassical reliabilism says is a necessary condition for objective justification: the property of being globally reliable under modulational control.

Consider, by way of contrast, a different agent who is also beset by a powerful deceptive demon (or computer). This agent, call her Faith, is provided with the appearance of a community that holds certain epistemic standards that are quite at odds with those that most of us have come to approve. For example, folk, or rather apparent folk, in Faith's epistemic community engage in the gambler's fallacy, and consistently approve of such inferences. They have no notion of the representativeness of samples, to take another example, and do not have evaluative practices that would lead to the associated caution in forming generalizations from instances. It is as if Faith has been raised in such a community—and has come to have the predictable inferential tendencies. Faith is not unreflective, however, and attempts to get at the truth as best she can. She conscientiously applies her epistemic standards—such as they are. (When bets or strategies informed by instances of the gambler's fallacy lead to disappointments or disaster, this is written off as cursed luck.) One day, Faith notes that it has been quite a while since the fair die used in a game has turned up a six. So, she forms the belief that a six is due—and that the probability of a six on the next toss is rather higher than 1/6—let us say she believes it is greater than 0.5. What is one to make of Faith's belief here?

Of Faith, one is no longer inclined to say (as one was of Constance), "Fine agent, lousy environment." One is inclined to judge that there is something about Faith's processes themselves (and not just the lousy environment) that makes them objectively inappropriate, and that makes her belief objectively unjustified.[8]

Of course, Faith's disposition to commit the gambler's fallacy is no less reliable in the demon-infested environment than were the processes that Constance employed.[9] Even so, one is inclined to judge that there is something wrong—objectively wrong—with Faith's processes, *and not with Constance's.* One is inclined to judge that Faith is not objectively justified in believing as she does, *although Constance is*. Constance is objectively and subjectively justified, while Faith seems only subjectively justified.

Apparently, the global reliability of belief-fixing processes is not a necessary condition for being objectively justified in believing. At least this is what seems indicated by one's judgments about Constance in her demon inhabited environment.

b. The "Truman Show" Scenario, with Harry and Ike

This scenario is structurally parallel to the Athena/Fortuna scenario, and elicits intuitive judgments suggesting that the property *being globally reliable as suitably modulated* is not constitutively required for justification. Although epistemic safety does indeed appear to be constitutively required for justification, an alternative account of this feature is needed. (Semantic externalist readers—who presumably doubt the intelligibility of the new evil demon scenarios—will find the scenario now discussed more to their liking.)

In the 1998 movie *The Truman Show*, the main character has been raised from childhood on a movie set where he has lived his entire life being the subject of a massive deception by skillful actors. The result has been a 24 hour a day television program in which viewers are treated to Truman's life from birth (which, one supposed, wasn't staged) to his being a young adult. We want to consider a competing show in which two friends, Harry and Ike, have been similarly deceived and observed. Harry and Ike spend their entire lives in carefully contrived circumstances in which everyone else they ever interact with is an actor playing a part in an elaborate conspiratorial deception. The conspiracy is so well orchestrated that it has rich "counterfactual depth": for various potential actions that Harry or Ike might engage in, including actions such as undertaking to change jobs or to travel to distant lands, the conspirators (the actors, directors, and stage-hands) stand ready and able to accommodate themselves in such a way as to maintain the ongoing deception. Thus, they are capable of doing their deceptive work within whatever alternative local environments Harry or Ike might venture into. Put simply, the deception of Harry and Ike extends potentially to such an extensive set of local environments that it is global.

Harry is highly paranoid, in one specific way. He believes that everyone else around him except Ike is an actor playing a part, and that every local environment he ever finds himself in is an elaborate stage-set being manipulated by designer/controllers behind the scenes. Harry possesses not a single shred of evidence for these sweeping paranoid beliefs, and he never obtains any evidence for them throughout the entire course of his life. Ike, on the other hand, exhibits no such paranoia. He takes his ongoing experience at face value, and has no tendency at all to believe or even suspect that he is being duped by a bunch of actors or that all his local environments are elaborate stage-sets. Ike believes that his friend Harry is just hopelessly paranoid, and he considers Harry's paranoid beliefs far too silly to take seriously for even a moment. (Ike too never receives even the slightest positive evidence in support of Harry's beliefs.)

We can say a little more about the etiology of Harry's curious belief forming processes. Harry is selectively given to certain kinds of fairly problematic generalizing practices in which he doesn't pay adequate attention to sample size and the like. He's only prone to get especially sloppy when certain emotions kick in: these emotions affect both his inductive-generalization processes and the wider processes that modulate them. Such emotions tend to kick in only on relatively few matters about which the sloppier version of the generalizing processes happen to be globally reliable—namely when he finds himself to have been the subject of pervasive deception. Given the skill of the actors/deceivers in his environment, recognition of pervasive deception would occur only when it was planned and intended by those deceivers. By plan, this occurred just once: and Harry became very upset upon having learned that his parents and all other adults were lying to him about Santa Claus. As a result, Harry's wider processes spawned or selectively triggered a process that has him perceiving deception everywhere (except when it comes to his friend Ike, who he (correctly) believes was also deceived). The triggered process is here globally reliable, as is the spawning process. Thus, the triggered process counts as globally reliable under modulational control—as this notion would seem to be understood by neoclassical reliabilists.[10]

Intuitively, however, Harry's paranoid beliefs are epistemically unjustified even so: they are highly *risky*, rather than being safe. Unlike in the case of Fortuna, this lack of safety cannot be understood in terms of a lack of global reliability in the processes that forms these beliefs (or in the modulating mechanisms that spawned those processes). The local reliability of Harry's processes does not depend on special or peculiar features of some particular local environment. All local environments are ones into which Harry and Ike's deceivers would readily precede them. So, all local environments are ones in which these agents would be subjects of deceptive presentations. For any potential local environment *L* within Harry's global environment, if Harry *were* to go into *L* then his processes *would be* reliable in *L*. The cumulative effect is that, inhabiting a global environment with able deceivers as his constant traveling companions, Harry's processes are globally reliable (as are the modulating mechanisms that spawned them). Still, the paranoiac belief forming processes strikes one as unacceptably risky or unsafe.

Recall that Fortuna's locally reliable processes seemed risky, despite being locally reliable, because that local reliability seemed dependent on peculiar aspects of her environment that her cognitive processes did not track and for which she had no informational basis, even no apparent informational basis. Similarly, Harry's processes seem risky, despite being globally reliable, because that *global* reliability is highly dependent on peculiar aspects of his *global* environment that his cognitive processes seem ill-suited to track, and for which he has no real indication. (Harry would have adopted the same paranoid beliefs and practices in the actual world, had he been the subject of a very common sort of Santa Clause deception as a child.) The comparisons and contrasts between the cases of Fortuna and Harry begin to come into better focus when one deploys the idea of a process that is yet more robustly reliable: a transglobally reliable process would not be highly dependent on peculiar aspects of the agent's global environment—because it would be reliable with respect to the class of experientially relevant possible global environments. It is this failure of the more robust form of reliability, in Harry's case, transglobal reliability, that makes for the unacceptable risk that we sense.

Ike's belief-forming processes fare differently. In the numerous cases in which his non-paranoid beliefs are at odds with Harry's paranoid ones, Ike's beliefs are systematically mistaken. But despite this very poor track record with respect to the goal of systematic true belief, intuitively these beliefs of Ike's are well justified nonetheless. This is because the processes that generate the non-paranoid beliefs are intuitively epistemically *safe*—even though they happen to be systematically non-veridical because of all that conspiratorial play-acting and behind-the-scenes set designing/controlling. As in the case of Athena, this epistemically laudable safety seems closely tied to the fact that Ike's belief forming processes here have a kind of robust reliability—but clearly, this is not to be understood as a matter of being globally reliable. Harry's processes are globally reliable (as are the modulational mechanisms to which they are attached), yet these processes lack the relevant epistemic safety, while it seems that Ike's processes lack global reliability and yet possess the relevant epistemic safety.

3. Lessons: Transglobal Reliabilism

For reasons of space, we must not continue proliferating scenarios. And we must quickly suggest some lessons. All the scenarios suggest that some notion of epistemic safety of belief-fixing processes is central to judgments deploying the concept of being objectively justified in believing. The judgments prompted by the cases of Athena, Fortuna, Diana, and Elena provide reason to believe that epistemic safety could be understood in terms of processes that are robustly reliable. Globally reliable processes are more robustly reliable than are merely locally reliable processes. So, this encouraged formulating and considering neoclassical reliabilism—which turns on global reliability under suitable modulational control. However, the new evil demon problem, in the persons of Constance and Faith, seems to indicate that such global reliability is not an adequate measure of the needed epistemic safety. Constance lacks it—and seems epistemically safe and objectively justified. She should count as better off from the point of view of objectively justified belief than Faith—although neither uses processes that are globally reliable. None of this suggests that the needed form of epistemic safety cannot be understood in terms of robustly reliable cognitive processes. Rather, it suggests that global reliability under suitable modulational control does not make for the required robustness of reliability. The cases of Ike and Harry do much the same work. Ike's processes are globally unreliable, but safe. Harry's are globally reliable under what the global reliabilist should think was suitable modulational control—but unsafe.

Things fall into place when one takes seriously the suggestion that robustness of reliability provides the needed key to what makes for epistemic safety and objective justification. As global reliability stands to local, so transglobal reliability stands to global—transglobal reliability is a more robust form of reliability than global. In all the scenarios discussed here, those who are judged to be justified in their beliefs deploy processes that are transglobally reliable—and those who are judged to be unjustified deploy processes that are not.

If we were to consider yet further cases, one would find that the concern for suitable modulational control needs to feature in transglobal reliabilism in a role parallel to the role envisioned in neoclassical reliabilism. Consideration of such cases must be reserved for another occasion. We can, however, express one general reason for thinking that modulational control remains significant. In the course of many investigations, one rightly fashions one's inquiry (including one's experiment and apparatus, if any) so as to be reliable in one's actual environment. A tailoring of the process of inquiry to one's local environment what was earlier noted in the cases of Diana and Elena. They each used information from *globally* reliable processes to condition their perceptual processes to a *local* environment, tailoring those processes so as to be there *locally* reliable. This modulational control enhanced the overall *global* reliability of their processes, so conditioned; thus we wrote of global reliability under suitable modulational control. In a parallel fashion, if one draws on *transglobally* reliable processes to inform and to provide modulational control of other processes, one tailors one's belief fixing processes so informed to one *global* environment in a way that is *transglobally* reliable. This enhances the *transglobal* reliability of one's cognitive processes. The epistemic payoff of having such control processes in place, of having various of one's belief fixing processes under the control of various transglobally reliable processes might be formulated in two complementary ways. First, the processes that one employs under such modulational control will *tend* to be appropriate to (reliable in) the possible global environment that one actually occupies, being tailored to that global environment in ways informed by transglobally reliable processes. Second, given that one's processes are under such modulational control, they would be reliable in a range of experientially possible global environments. Having one's belief-fixing processes under the modulational control of transglobally reliable processes makes for the transglobally reliability of the package of processes one employs. This clearly contributes in desirable ways to the epistemic safety of one's cognitive processes, and foregoing such modulational control is ceteris paribus unnecessarily risky. So, ultimately, we suggest that what makes for objective epistemic justification is the use of processes that are transglobally reliable under suitable, transglobally reliable, modulational control.

It is worth noting that degrees of robustness of reliability, and now of transglobal reliability, can again be understood by thinking of what variations across environments would preserve reliability. As we stressed earlier, when one compares a globally reliable and a merely locally reliable process, one can begin by thinking of a local environment in which each is reliable. One finds that the merely locally reliable process would fail to be locally reliable in any but highly similar local environments, while the globally reliable process would remain locally reliable across more variations in local environment. Similarly, one might compare two processes: one of which is transglobally reliable, and the other being merely globally reliable. The reliability of the merely globally reliable process is dependent on the particulars of that environment in such a way that it would fail to be reliable in any but highly similar environments. The global reliability of a transglobally reliable process is not so sensitive—and would be retained across greater variations in global environment. To provide a quick and relatively simple illustration, consider two inductive processes. One incorporating sensitivity to sample bias, size, and the like, while the other is without such sensitivity. Then imagine a global environment with sufficient homogeneity of populations and causal structure that both processes would be globally reliable. The first is transglobally reliable, while the second is merely globally reliable. As one imagines global environments with increasing heterogeneity of populations and causal structure, one quickly comes to global environments in which the inductive processes without sensitivity to sample properties becomes globally unreliable—while the processes incorporating sensitivity to sample properties would continue to be globally reliable. The transglobal reliable process is the safer.

The new evil demon problem gives expression to concerns that are commonly taken to be important by internalist epistemologists. As we have seen, transglobal reliabilism accommodates these concerns or scenarios easily. With its demand for transglobal reliability under suitable modulational control, it would have no difficulty also accommodating BonJour's clairevoyant scenarios. Yet, transglobal reliabilism is an externalist epistemology. Its continuity with more familiar forms of externalist reliabilism can be appreciated in terms of its central motivation. Thus we have here emphasized the concern for epistemic safety and the sense one has that epistemic safety is intimately connected with the robust reliability of one's belief-fixing processes. We have argued that epistemic safety is best understood in terms of adequate robustness of reliability under suitable modulational control. There is another continuity with famliar forms of reliabilism—one that also contributes to the externalist character of transglobal reliabilism: transglobal reliablism, like reliabilisms generally, is concerned that agents employ or deploy processes that are appropriate to the kinds of cognitive critters they are. This is to think in terms of processes that are tractable for a class of cognitive systems—notably humans, who constitute the standard focus of our epistemology. The concern for tractible processes has remained largely implicit in the foregoing, but it is clearly there, in the demand for processes that are "adequately transglobally reliable *under suitable modulational control*." For, what control processes are suitable for a given class of agents cannot be understood independent of an understanding of their cognitive architecture, the degree of plasticity of various of their cognitive processes, and the extent to which they can (with training and motivation) acquire and effectively deploy various processes. The concern is that the agent acquire and deploy belief-fixing processes that are satisfactorily transglobally reliable under modulational control, and processes that, taken together, are among the more transglobally reliable sets of processes that such agents might tractibly employ. What is particularly striking is the extent to which this externalist demand readily and naturally lead one to accommodate concerns that have been associated with internalist epistemologies.

References

BonJour, L. 1985: *The Structure of Empirical Knowledge*, Harvard University Press.

Henderson, D. and Horgan, T. In press: "The Ins and Outs of Transglobal Relativism", in S. Goldberg (ed.), *Internalism and Externalism in Semantics and Epistemology,* Oxford University Press.

Henderson D. and Horgan, T., 2001: "Practicing Safe Epistemology," *Philosophical Studies* 102: 227-58.

Lehrer K. and Cohen, S. 1985: "Reliability and Justification," *The Monist* 68: 159-74.

Moser, ()

Plantinga, A. 1993a: *Warrant: The Current Debate*. Oxford University Press.

Sosa, E. 1991. "Reliabilism and Intellectual Virtue," in E. Sosa, *Knowledge in Context,* Cambridge University Press, pp. 131-45.

[1] We discuss the internalist and externalist elements to transglobal reliabilism in Henderson and Horgan (in press). The seeds of transglobal relilabilism are found in Henderson and Horgan (2001).

[2] Note well: A possible global environment can be compatible with one's having such experiences within it even if these experiences are radically and systematically nonveridical.

[3] We are talking here about employing these processes indiscriminately,without possessing information to the effect that one is presently in one of the pertinently unusual environments. The more selective use of such processes is a matter we address in section 2.1.f below.

[4] It might be tempting to think that the reliability of modulated processes can be no more robust than is the reliability of modulating processes—and if the modulating process is merely locally reliable, than so must be the modulated processes, as modulated. While this is generally correct, there are special cases. Suppose, for example that process *P* serves to inhibit process *Q* when *P* registers certain circumstances, *C*, that would make for *Q*'s being locally unreliable. Now suppose that *P* itself is not globally reliable because it is given to generating too many false positives regarding the presence of circumstances C, although given circumstances *C, P* is locally reliable regarding *C* (it does not give false negatives). Here, *P* might provide a form of modulational control with respect to *Q* that makes for a tailored employment that is more globally reliable than it *Q* would be without such control.

[5] At least this holds for empirical beliefs that the agent might generate. The case of *a priori* beliefs, particularly high-grade *a priori* beliefs is not addressed here. Also, we are here assuming that empirical beliefs pervasively of the form "The Deceiver is providing me excellent evidence for the false statement that …" are non-starters as candidates for being objectively justified (even though they happen to be true), since agents in the envisioned scenario have not a shred of evidence for such radically paranoid empirical beliefs.

[6] Of course, because she is subject to demonic deception, her training has of necessity been reflective of false-successes and false-failures. But, otherwise it reflects the sort of training and shaping of perceptual processes that makes for systematic success in our global environment.

[7] Note well: One judges that Constance's beliefs are *objectively* justified—not merely that they are subjectively justified relative to her own epistemic standards. Her epistemic standards are objectively just fine, even though she's in a damn lousy global environment.

[8] Faith may be *subjectively* justified in believing, and one tends to evaluative her conscientiously formed beliefs accordingly. Given her observation of a long enough run of non-six tosses, and her beliefs or standards, she is subjectively justified in her belief that a six is due. Still, one is far less inclined to judge that Faith is objectively justified.

[9] Constance's dispositions are better than Faith's as a basis for predicting the future course of first-person experience, of course. But Constance's are just as bad as Faith's with respect to the truth values of the external-world beliefs they generate.

[10] If the triggered process had been generated by a globally unreliable modulating process, then the neoclassical reliabilist could say that the paranoid beliefs produced by the triggered process are not well justified—even though the globally unreliable modulating process happened to spawn, in Harry's case, a paranoid belief-forming process that itself just happens to be globally reliable. But the key point to appreciate is this: In Harry's case, the modulating process *itself* happens to be so structured that that it is disposed to trigger only certain quite specific forms of shoddy inductive reasoning—forms of reasoning that happen to be globally reliable in Harry's specific global environment. Thus, Harry's paranoid belief-forming process was triggered by a globally reliable modulating process.