UNIVERSITY OF CALGARY

JUSTIFICATION, RELIABILISM AND THE PARSING PROBLEM

by

Allen N. Habib

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTERS OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

CALGARY, ALBERTA

AUGUST, 1999

Abstract

There has been a long-standing tradition among epistemologists to charcterise justification solely in terms of the evidence a person has in support of a given belief. This evidentialism has serious drawbacks, however, and around the middle of this century a new movement arose, one that eschewed the traditional 'internal' limits on the framing of justification, and looked towards other factors. The most successful of these 'externalist' theories was called process-reliabilism, and it held that the justification of a given belief was a matter of the reliability of the mental process that led to the belief's being formed. But this approach also suffers from severe problems, notably the difficulty in parsing mental processes and the possibly false and certainly not scientifically motivated picture of cognitive psychology it requires. I propose a variation on the theme of process-reliabilism, one wherein justification is measured by the performance of the entire agent in epistemically relevant circumstances, rather than the performance of the individual mental process of the agent's that gave rise to the belief. This framing alleviates the problems in parsing the putative mental entities, while retaining the strengths of an externalistic and naturalistic theory of justification.

Acknowledgements

I would like to extend my sincerest thanks to all the people who helped me during the research, writing and editing of this manuscript. Specifically, Dr. Ann Levey, my thesis supervisor, and Dr. Steven DeHaven, my principal outside reader and editor. Without these two, this work would simply not have been attempted, at least not by me. I would also like to gratefully acknowledge the invaluable assistance I received from the faculty and graduate students at the University of Calgary, and at Syracuse University.

TABLE OF CONTENTS

Chapter One
Justification and reliabilism, an overview

1.1 A brief history of justification

It is generally agreed that for a person S to know a proposition P, S must believe that P, and P must be true.[1] But this is not the whole story. Truth and belief, while individually necessary, are not jointly sufficient for knowledge, as Plato demonstrated in the Theatetus.[2] To rehearse just one of his arguments here: if a person guesses something to be true, and, fortuitously, it is true, that person has a true belief, but that belief is not a piece of knowledge for him, it is not something that he knows. So what is the difference between knowledge and mere true belief? In order to account for this difference on an analytic framework at least one other condition of knowledge is necessary. This third condition is commonly called justification.

Prior to the 1960's, the standard view of justification was evidentialist. Going back to at least Descartes, epistemologists in the main believed that the justification for a belief, that which separated what someone knows from lucky guesses and other 'merely true' beliefs, was the evidence a person had for holding that belief.[3] For an evidentialist a belief is justified just in case the believer holds sufficient evidence of its truth to warrant her holding it. On such a picture, justification is epistemically internal to the believer, meaning that the justification (i.e., the evidence or reasons) for any given belief is accessible to the mind of the believer, it is something she could come to be aware of upon reflection alone.[4]

There is no question that this internalist, evidentialist view of justification is appealing. The rules of evidence certainly do seem pertinent to the concept of justification.

To see this, consider two demon-world cases. In the first case Sally has her senses consistently deceived by a demon, so that the beliefs that she forms on their basis are all false, but Sally reasons correctly from these false premises, and forms new beliefs only when she has adequate evidence for doing so. Now contrast Sally with Frank, who is also bombarded with false information, but who eschews the rules of evidence altogether, and forms his new beliefs entirely capriciously. Obviously Sally has something over Frank, in an epistemically important way. An evidentialist would argue that Sally is justified in her inferential beliefs *because* she follows the rules of evidence, whereas Frank, who does not, lacks justification for his beliefs, and this is why she seems epistemically superior. This is not implausible. Certainly Sally's adherence to the canons of inference seems to make her beliefs more justified than Frank's.

But evidentialist views of justification suffer perennial problems. One notable problem is that, on such a view, the justification a person has for a belief consists solely in other beliefs of that person, beliefs that serve as the evidence the person has for the target belief. But if this is the only method of justification, then those justifying beliefs must themselves be justified by still further beliefs of the person, since it would not behoove us to include unjustified beliefs in our justificatory arguments. Whatever a piece of evidence is exactly, it must at least be something that a believer believes justifiably.[5] But then those further beliefs would themselves have to be justified by appeals to still other beliefs, and so on *ad infinitum*. Thus the evidentialist finds herself in a dilemma: either she must posit that agents can have an infinite number of beliefs, which seems impossible on its face, or she must admit circularity in all justificatory arguments.

Often called the problem of the infinite regress of justification, this dilemma is endemic to evidentialism. Two different methods of stopping the regress cleave evidentialists into two broad camps, foundationalists and coherentists. Foundationalists propose to stop the regress by positing 'basic' or 'foundational' beliefs. These are beliefs that derive their justification from sources other than other beliefs of the agent. Foundational beliefs are often held to be self-justified, either because of their provenance as sensory or apparent beliefs or from their status as logical or pragmatic truths. Regardless of how foundational beliefs are justified, what makes the theories that posit them foundational is that the regress of justification is halted by arrival at one or more of these beliefs. Upon arriving at a foundational belief in a justificatory argument, we need no longer seek further beliefs to justify it, thus halting the regress.[6]

Coherentists stop the regress by falling on the second horn of the dilemma and admitting circularity in justificatory arguments. For a coherentist, the justification of a belief lies in its 'coherence' with the other beliefs of the person, where coherence is usually explained in terms of some logico-semantic accordance with other beliefs. For coherentists justificatory arguments are not free-standing chains of inference ending (or starting with) self-justified basic beliefs, but rather smaller parts of a self-supporting web of beliefs, and it is the fact that they 'fit' into such a web that justifies individual beliefs.

Both methods of solving the regress suffer intractable problems. To pick just two as a sample, foundationalists are at a loss to find suitable candidates for the beliefs they propose as foundational,[7] and coherentists are at pains to explain how sensory and other somatic and apparent beliefs are justified, since they are justifiably believed even in cases

where they do not cohere with the web of beliefs.[8] As well, in 1963, Edmund Gettier published a paper outlining another problem for evidentialist views.[9] By use of ingenious thought experiments, Gettier demonstrated that the conditions for knowledge of an evidentialist theory might be met, yet the belief still not be knowledge for the believer, due to the accidental nature of the truth of the belief. The Gettier problem seemed intractable both from foundational and coherentist points of view, and in this climate a group of new theories of justification emerged.

A common thread running through most of these new theories was the conception of justification as a link between the truth of a belief and the reason that a believer holds it. For this reason I will call them truth-oriented theories. The different theories spell out the link between truth and belief in a variety of ways. D. M. Armstrong, for example, proposes a causal/nomological link, such that a belief would be justified if its truth was the cause of, or nomologically connected to, its presence in the mind of the believer.[10] Truth-oriented theories are commonly externalist, meaning they do not require that a believer have privileged access to the justification of her beliefs. They are also quite commonly naturalistic theories. Most externalist theories peg justification to some aspect(s) of the physical world.

I believe that one of these new externalist, truth-oriented theories, reliabilism, is the most promising candidate for a complete and correct explication of the justification concept. Reliabilism was popularised and refined by Alvin Goldman in his papers and books beginning in the early 1960s[11], and saw its most detailed explication in his 1986 book Epistemology and Cognition (E&C).[12] I also believe, however, that Goldman's

theory suffers from a number of conceptual difficulties, difficulties that mandate substantive emendations to the theory. My aim is to provide and defend such a revised theory in this thesis.

1.2 Process Reliabilism, the theory and its problems

The most general framing of the reliabilist position would be something like this: 'Justification is a matter of the reliability of the causal mechanism that produces a belief. S is justified in believing that P if whatever caused P to be held by S was a reliable causal process'. 'Reliable' here means truth-productive, so a causal mechanism is reliable if, over a series of actual or counter factual belief productions, some specified percentage of the issuing beliefs are true.

One of Goldman's earliest refinements to his reliabilist theory of justification was his proposal to limit the pertinent causal mechanisms solely to the cognitive processes of the believer in question.[13] In Goldman's theory, from 1976 onward, the justification for a belief is a matter of the statistical measure of the truth-production ability of the *cognitive process* that caused the belief, and a belief is justified just in case its Belief Forming Process (BFP) is one that meets some specified ratio of truth-attainment.

Goldman calls this theory process-reliabilism (PR), and it has enjoyed no small measure of success among contemporary epistemologists. It successfully handles such epistemic trouble-spots as the justification of sensory and other non-reflective beliefs, and the ascription of knowledge to children and animals. As well, not being evidentialist it does not suffer from the regress problem.

PR is externalist because the justification-conferring properties it proposes,

namely the truth-attainment ratios of the issuing BFPs, are not necessarily internal in an epistemic sense to the believer. The reliability rating of my cognitive processes is not something that is 'available' to me in any epistemically meaningful sense. And PR is truth-oriented in that it proposes a statistical link between the truth of a belief and the fact that it is held. It is the ratio of true beliefs to false of a belief's productive BFP that justifies it.

But process-reliabilism is vulnerable to a variety of criticisms. One charge is that the condition of the reliability of the generating process, when combined with truth and belief is insufficient for knowledge. Laurence Bonjour makes this argument in his 1981 article 'Externalist Theories of Empirical Knowledge'.[14] He argues that in cases where a person believes something as the result of a reliable process, but has other beliefs that conflict (in some epistemic sense) with the reliably issued belief, that person might not know the belief in question, regardless of its truth or of the reliability of its issuing process; he eloquently summarises this point by saying: "Objective reliability is insufficient to offset subjective irrationality."[15]

In support of this argument, Bonjour gives the following example: A true belief about the whereabouts of the U.S. President is formed in Maude by a reliable process of clairvoyance. But Maude has overwhelming evidence, in the form of TV and newspaper reports that the President is in another location than the one her clairvoyant power indicates. In this case, even if Maude's belief were true, she cannot be said to know the President's location.

But, in these circumstances, any plausible justification condition generated by process reliabilism is satisfied. The process that formed the belief in question is *ex*

hypothesi reliable to any degree the reliabilist would care to set. Thus process-reliabilism does not generate a sufficient condition of justification. This I call "the problem of defeated justification".

Another argument that tells against PR is the problematic nature of the central notion of a 'process', specifically, the notion of a Belief-Forming-Process. Problems with the notion arise in two ways. One is an argument first introduced by Richard Feldman in a 1986 article entitled "Reliability and Justification"[16], and echoed by Plantinga[17], among others. It goes as follows: Goldman claims that a statistical property, like that of having a certain truth-attainment ratio, can only be properly said to be possessed by a type of thing, rather than a token. But, the objection goes, What must then be measured for its truth-ratio is a type of belief forming process, a BFP-type, as opposed to a token BFP, or an actual physical instance of belief formation. But an individual BFP can be a token of any number of types of BFP. For example, S's belief that she is Napoleon might be the result of BFPs that are, respectively, tokens of the BFP-type 'processes terminating in beliefs about S's being Napoleon' and of the BFP-type 'cognitive processes occurring in S'. The former type is obviously not reliable, whereas the latter might well be. How do we choose the relevant BFP-type? Presumably, only one of the BFP-types that a token BFP falls under is relevant to the measure of reliability, at least as far as justification-conferring reliability is concerned, so how shall we decide which one that is. It is incumbent upon the reliabilist to provide criteria for determining which BFP-type is the correct one.

Here the reliabilist might point to Goldman's proposal that only the 'cognitive processes' of a person be considered as parts of a BFP, and argue that it greatly narrows

the number of potential BFP-types to choose from. But even granting this, Feldman argues that the notions of BFPs and BFP-types are still problematic, because criteria for the proper 'level of generality' are still needed. Were BFP-types to be defined too specifically, Feldman argues, they would contain only a very few or a unique token(s), and as such would not be proper candidates for statistical measure. As well, they would not be useful as ways of explaining the concept of justification, since they would effectively collapse justification to truth. Similarly, if process-types were to be defined too generally, then they would be equally useless as objects of statistical measure, at least for the purposes of determining justification, since one process-type would account for an enormous number of beliefs, some of which would have obviously different justificatory statuses.

For instance, if we define vision as a BFP-type, and take the statistical measure of the truth ratio of all the vision-beliefs of a person as the determination of the justification of a given vision-belief of that person, then we are stuck saying that a person's vision beliefs all have the same degree of justification. But imagine a case of defeated justification. Say an agent comes to believe that a shirt in her field of vision is blue, while she is wearing blue-tinted lenses. Further suppose that at the time that she forms this belief, she is aware that she is wearing blue-tinted lenses, and aware that blue-tinted lenses have the effect of making (non-blue) things appear blue. It certainly seems that the agent would be unjustified in her belief because of her inattention to these facts. However, on a PR schema where the relevant BFP-type is vision, she is justified in her blue shirt belief just in case some high enough percentage of her *overall* vision beliefs are true. But

obviously here the person should have taken into account the fact that things appear blue to one when one is wearing blue-tinted lenses, and her failure to do so removes her justification for the blue belief, regardless of how good she is at coming to true vision beliefs in the main.

The task of framing BFP-types such that they are neither too broad nor too narrow is what Feldman calls 'The Problem of Generality'. It is the problem of giving criteria for the definition of process-types that result in types that are neither too wide nor too narrow. I lump all of these problems together under the heading of "the parsing problem".

The second problem with the notion of a belief-forming-process is the following. Goldman says that the most plausible way to specify BFP's is to make them sets of psychological events in the believer. For Goldman, BFPs and BFP-types are psychological entities. What follows from this is that, were BFPs and their types actually to exist, they would be entities of great interest to cognitive scientists. Indeed, if BFPs (or BFP-types) were actually to play the pivotal role in justification that the process-reliabilist accords them, as the objects whose statistical measurements determine the justificatory status of a belief, then they would be psychological objects of crucial epistemic importance, and this fact alone would be ample motive for their scientific study.

But this is exactly the problem. What evidence do we have that these objects exist? The psychological casting of BFPs and their types commits the process-reliabilist to a rather detailed theory of cognition, and to a psychological ontology. PR forces its proponents to adopt a detailed story about how the mind works. Worse still, the entities described in the theory are not discovered by the observational methods of scientific

inquiry, but rather they are (or they seem to be) delineated by the epistemic demands of a theory of justification. This amounts to philosophers dictating to scientists as to the nature of the empirical world, and such projects are notoriously unsuccessful. Any theory of knowledge or justification that turns on such a demand for a detailed theory of psychology is jumping the gun, as it were, putting the theoretical cart before the scientific horse. As such, the charge goes, process-reliabilism, because of its reliance upon processes as putative objects of psychological reality, is unacceptable. These two problems, the problem of generality and the problem of psychological reality, I unite under the heading of "the problems with the notion of a belief-forming-process".

These two types of problems, the insufficiency of the reliability condition and the unacceptableness of the notion of a BFP, are, I believe, fatal to process-reliabilism. But can some version of reliabilism succeed? I think that the answer to this is yes. Reliabilism contains insights into the nature of justification that I believe are substantially correct. These are: 1) That justification is a matter of cognitive performance, 2) that the measure of justification is the measure of an agent's cognitive performance in a given scenario, and 3) that the correct standard of performance measurement is truth attainment. In chapter four, I offer an amended theory of justification that retains these insights while avoiding the two difficulties outlined above.

1.3 A new proposal, System Reliabilism

To review, the problems with the process-reliabilist approach are that: 1) The measure of the reliability of individual BFP-types is an insufficient method for determining a cogniser's overall justification for a given belief, since a belief might be reliably produced

in the believer, but the believer might have good reasons to disbelieve it, and 2) that the notions of BFPs and BFP-types are problematic because the criteria for their proper delineation is lacking, and because, on such an account, these processes are supposedly real psychological entities, but they lack scientific backing.

In response to these criticisms, I propose a version of reliabilism that results from changing the objects of cognitive measurement from processes to systems, but not the standards of their measurement. The core idea of the theory is that it is the performance of the cogniser as a whole, as a cognitive system, that is crucial to justification, and not just the performance of the cogniser's BFP-type, or whatever the actual component of the cogniser's cognitive apparatus it was that gave rise to the target belief. I call the theory system-reliabilism, or SR. In SR, an agent S is justified in her belief that P so long as S, as a cognitive system, is reliable enough at forming P-type beliefs to be justified in believing that P. SR eschews all talk of the psychological entities within a cogniser, and instead focuses on a cogniser's epistemic behavior in a variety of (epistemically relevant) circumstances. As in process-reliabilism, the standard for measuring cognitive performance is truth attainment, but in SR what is measured is not the truth-attainment-ratio of a part of the cogniser's mental apparatus, such as a certain BFP-type, but rather the truth-attainment ability of the system (the cogniser) as a whole, with regards to situations that are epistemically relevant to the belief in question.

In SR the method of justification measurement is the determination of truth-attainment in a variety of possible or actual situations. These situations form a set, which I call the "test-set", for the belief in question, and the judgment of reliability (and hence

justification) is based on the truth attainment ratio of the system among the members of this set. The selection of a test-set of belief/believer situations is determined via a criterion of epistemic relevancy. When is one situation relevant to another, epistemically relevant, such that the performance of the cogniser in that situation will go towards determining the cogniser's reliability, and hence her justification, for the belief at hand? I propose a conception of epistemic relevancy that turns on the notion of capacities that underlie our ability to make accurate judgements. For SR, a scenario is epistemically relevant to a target belief if that scenario embodies conditions that test the capacities that our best scientific knowledge tells us are necessary for forming an accurate target belief. Of course, our societal codes and practices will also play a role in determining epistemic relevancy.

1.4 The benefits of System Reliabilism

What we gain by making this move from process to system reliabilism is the retention of the reliabilist insights into the nature of justification, as well as PR's ability to handle the aforementioned epistemic problems, while at the same time being able to answer the problem of defeated justification, and avoiding the problematic notions of BFPs and BFP-types altogether. To handle defeated justification on SR we need only point out that cases where a believer has negative evidence (or a lack of any positive evidence) for the type of belief in question are cases that test for capacities (say, sensitivity to the truth) that the believer should have to be justified in the target belief. Thus situations of this sort are candidates for the test-set of the target belief As such, if an agent performs poorly on these scenarios, she will fail her justification test, and thus not know her belief, a result that coincides with our pre-theoretic judgement.

So, for example, in our earlier case of a cogniser coming to believe that something in her visual field was blue while wearing blue-tinted lenses, on an SR account we would say that situations like that one, situations embodying defeating conditions for the type of belief in question, would be situations that would appear on the test-set for this particular belief. If our cogniser does as poorly on this section of the test as she does in the instance given, then she will surely fail, and as a result be unjustified in her belief

And of course, since the amended theory does not require the positing of BFPs or BFP-types, the problems involved in defining such entities, as well as the problem of positing such empirically unsupported entities at all, are avoided. The parsing problem, however, must be addressed, since SR also contains some theoretical constructs (belief-scenarios, test-sets) that might well suffer from analogues of the generality problem and the single case / no distinction dilemma.

In what follows I will present and defend this revised version of reliabilism. In the next two chapters I will give a more detailed account of Goldman's process-reliabilism, and examine his and other's attempts to respond to the problems I have outlined for a process-reliabilistic framework. I will show how and why the process arguments are problematic, and in the fourth chapter I will provide a detailed account of system-reliabilism. In the final chapter I will demonstrate how SR can handle the problems that stymie PR. As well I will consider possible objections to SR, and then provide general arguments in favour of externalist and reliabilist views of justification, and specific arguments in favour of an SR framing in particular.

Chapter Two

Process Reliabilism and Defeated Justification

2.1 The Problem of defeated justification

Goldman puts forth a very sophisticated version of process reliabilism in E&C, one that contains a number of amendments to previous formulations designed to handle the types of problems outlined above. In this chapter I will consider his attempts to solve the problem of defeated justification. First I will examine the types of cases that give rise to the problem, and then turn to Goldman's responses. Then I will detail the problems with Goldman's methods of handling these cases.

There are a number of ways that cases of defeated justification can be classified. I will divide the cases into two broad types, negligent and non-negligent. Non-negligent cases of defeat are cases, like Gettier cases, where the belief's justification is removed through no fault of the believer. Negligent defeat cases are those where the cogniser does bear some responsibility for their belief's lack of justification. The former group I will return to in the final chapter. For now, I will focus on the latter.

I will categorise cases of negligent defeat into three sub-types. These are: 1) Positive defeat cases. These are cases where a cogniser has evidence, in the form of other beliefs or mental states, which undercut or rebut the belief in question. 2) cases of negative defeat, where a cogniser lacks sufficient evidence, in the form of other beliefs or mental states, to adequately support her belief, and 3) cases of simple defeat. These are cases

where a cogniser has no positive defeaters, and is in possession of sufficient evidence for their belief, but she lacks beliefs that she ought to have in order to be justified in the target belief. In these last cases some epistemic failure on the part of the agent, past or present, defeats her justification for a belief, regardless of the coherence of her belief-corpus at the time of the belief's formation.

Laurence Bonjour's cases involving clairvoyants[1] are examples of the first two types of negligent defeat cases. In Bonjour's case III, Maude has a reliable clairvoyant power, and believes that she does, but she has overwhelming evidence that such a power does not exist. She forms a belief as to the whereabouts of the U.S. President as a result of this power, and the belief is true, but we do not want to say that she knows where the President is. In case II, Casper also believes (truly) that he is clairvoyant, and he is also in possession of massive amounts of evidence to the contrary. His evidence is in the form of inductive conclusion from his numerous failed attempts to demonstrate his clairvoyant abilities. He also comes to a true belief regarding the President's location as a result of his power, but again, we feel he does not know this. In Bonjour's case I, Samantha, who is similar to Maude and Casper in that she is clairvoyant and believes so, comes to a belief about the location of the president in the face of massive evidence to the contrary, but in this case, we are not told about her evidence for her belief in her clairvoyant power. Rather the contrary evidence is in the form of news reports that she has seen, reports that she has every reason to belief trustworthy, that tell her the president is somewhere other than where she believes he is.

All of these cases are examples of what I call positive defeat. Both Maude and

Casper ought not to believe what they do (and are as a result unjustified and thus lack knowledge) because they are in possession of other beliefs that undercut the evidence for the target belief. These beliefs make the target belief unjustified because they cast doubt upon the manner in which the belief was formed. Samantha is in possession of a set of rebutting[2] defeaters. These are beliefs that make the target belief unjustified because they are logically incompatible with it. What these examples demonstrate is that failure to take into account either type of contrary evidence on the part of a cogniser can eliminate the justification for a given belief.

Bonjour also presents a fourth case that demonstrates what I have called negative defeat. In case IV Norman possesses a reliable clairvoyant power, but does not believe himself to be clairvoyant. Indeed he has no beliefs about this fact, nor about the existence of clairvoyance in general. Norman also comes to a true belief about the President's whereabouts by dint of his power, but again, we are very reluctant to say that he knows or is justified in this belief. The reason for this reluctance is that Norman has no reason to believe what suddenly pops into his head. Even without positive evidence to the contrary for a belief, we feel that a cogniser ought not to believe something in the absence of some good reason to believe it.

As for simple defeat, in E&C Goldman provides two examples, one by Putnam and one of his own.[3] In Putnam's hypothetical case a person, say Bob, comes to believe that the Dalai Lama is infallible on matters of faith and morals. This turns out to be the case, and, as a result, Bob has hit upon a perfectly reliable method of making judgments about ethical matters. But we feel that Bob does not know the beliefs issuing from the cleric to

be true. Bob is unjustified in his true beliefs about morals, because of the bizarre way he came to adopt them. Goldman's own example of negligent defeat is that of Humperdink, a gullible math student, who learns a perfectly correct algorithm from Elmer Fraud, whom Humperdink has been warned is no authority on mathematics. Regardless of the truth of the beliefs formed in Humperdink as a result of listening to Fraud, Humperdink does not know the algorithm.

Upon closer inspection these two examples might seem better categorised either as cases of positive or of negative defeat, depending on what information we get about the noetic structures of the cognisers at the time of the formation of their respective beliefs. If Putnam's acolyte believed, for example, that no one is infallible, prior to forming his belief about the Dalai Lama, then we might say that he has a rebutting defeater for that belief, and consequently is unjustified for this reason in his belief about the cleric's infallibility, and so unjustified in beliefs that follow from this belief. Alternately, if Bob believed the Dalai Lama infallible because the belief simply popped into his head, and he was not in possession of any evidence for it save its presence, then we might say that it is a case of negative defeat, and that Bob's ethical beliefs are unjustified because they rest on such a negatively defeated belief.

Humperdink, in Goldman's example, comes to believe what Fraud tells him despite having "been warned" about Fraud's untrustworthiness. This certainly seems a case of positive internal defeat of the undercutting kind, and we can envisage an analogous case where Humperdink is the victim of both positive rebutting or negative defeat, depending on the beliefs we assign him.

So both of Goldman's examples of negligent defeat seem unsatisfying, since they both seem more easily characterised as cases of positive or negative defeat. This might lead us to question the viability of the third category altogether. Perhaps all cases of simple negligence are, like these two, only apparently members of the third group, and perhaps any supposed member of this group will be found to be a member of one of the other two groups (positive and negative) following a sufficiently thorough investigation of the cogniser's belief corpus. I am not of this mind. I believe that there are cases of negligent defeat that cannot be accommodated under the definition of the first two types. Cases of simple epistemic defeat are cases where justification is defeated without the presence of positive or negative defeating evidence in the believer's mind. Goldman realises that this category is not empty, although his examples aren't convincing, and as we shall see, he formulates a complex stratagem to handle this type of case. But we first should provide an unambiguous example:

George attends classes at a local college, and every week a visiting lecturer takes the place of his philosophy professor to give a guest lecture. There is always an introduction of the guest on the part of the professor, and George always studiously disregards it. This is because of a peculiar belief of George's, borne out by induction from an unfortunate string of happenstances in George's life, that professors are not to be trusted in the least in matters of information concerning other professors. One day the professor introduces a colleague from the psychology department, and tells the class that, as part of an experiment, the guest is going to tell them clever half-truths, mixed in with actual facts, to see if they can spot the difference. As it turns out, this is not the whole

truth. There is an experiment in progress, but its purpose is to determine educated people's reactions to truths when they have been prompted to disbelieve their source. Towards that end, the psychologist will only tell them the truth, after the disingenuous introduction by the philosopher. George is, of course, ignorant of all of this, and when he tunes in to the psychologist, he forms beliefs about what is being said, and he takes these beliefs to be true.

Does George know the truths presented? I believe the answer is no. George does not know all of the truths uttered by the psychologist, because had he not had his particular epistemic quirk, he would have paid attention to the introduction, and as a result he would not have believed all of the truths recounted. What is important to remember here is that other than for his peccadillo concerning academic introductions, George is epistemically exemplary, and that at the time of his new belief's formation his noetic structure was adequate to justify them. George has no evidence that the beliefs he forms are not true, and he has good reasons to believe that they are, since they are being issued by an accredited expert in a scholarly and sober setting. So George suffers from neither positive nor negative defeat. The only plausible reason we have to deny George knowledge (and hence justification, since his beliefs are true) is his anterior decision to disregard academic introductions. It is this past failure that causes him to be unjustified in his beliefs, not his present belief corpus, and this past failure, as this example demonstrates, does not automatically lead to the presence of positive or negative defeat conditions for the target belief in George's mind.

2.2 The non-undermining notion

So how does Goldman account for these three cases of defeated justification? Firstly, Goldman strengthens his framework (formal) justification condition, which initially reads:

> "(P1) S's believing that P at t is justified if and only if it is permitted by a right system of J[Justification]-rules."

by adding a further restriction, that the belief in question not be what he calls 'undermined' by the cogniser's mental state at the time of its formation, time t. This yields P3, his actual formal justification condition:

> "(P3) (a) S's believing that P at t is justified if and only if it is permitted by a right system of J-rules, and
> (b) this permission is not undermined by S's cognitive state at t."[4]

The obvious question is 'what does it mean for a belief to be undermined by a person's cognitive state? Goldman lists three types of undermining cases. First, there cases where a believer believes that the target belief is not permitted by a right set of J-rules. Then there are cases where a believer is permitted to believe that the target belief is not permitted by a right set of J-rules, regardless of whether or not they actually believe so. And finally there are cases where a believer believes that certain conditions that are, in fact, necessary for a belief's permissibility do not obtain, regardless of whether the believer believes these conditions to be necessary for the belief's permissibility. These three types of undermining are roughly equivalent to the first two types of defeat cases that I outlined above. Take the example Goldman himself supplies of Millicent.[5] Millicent is in possession of cogent reasons to distrust her eyes, since a qualified expert has told her that she has ingested a powerful visual hallucinogen. However Millicent ignores this warning and continues to form and hold visual beliefs in the normal manner. As a result she is

unjustified in her beliefs. This is obviously a case of Millicent being in possession of undercutting evidence about her eyes that defeats the justification she has for her visual beliefs.

And indeed Goldman uses this non-undermining condition to deal with the types of defeat cases introduced by Bonjour,[6] expanding his definition of undermining to account for the other types of defeat that Bonjour's cases exemplify, namely negative defeat. Goldman does so by positing a theory of what he calls *ex ante* justification, a theory that deals not only with what the cogniser does believe, but also with what she would be justified in believing, regardless of whether or not she does believe so.[7] Goldman proposes that his treatment of justification thus far could easily be developed to include this type of justification, with at least one important modification, namely the inclusion of J-rules of obligation in addition to those of permission. Such rules would require that cognisers take into account rebutting as well as undercutting defeaters for their beliefs in order to be justified.[8]

What this means for cases like those of Casper's and Maude's discussed above is that Goldman can say that their beliefs' permittedness is undermined because they ought to believe that it is not permitted, and this obligation stems from a J-rule that says, roughly, the cogniser must take into account contrary evidence. The case of Norman, which exemplifies negative defeat, Goldman also handles by use of the non-undermining clause. He does so by proposing that certain obligation rules are necessary for any adequate J-rule system, at least for any that would confer justification on a normal human cogniser. It is difficult to envisage, says Goldman, such a normal cogniser coming to

believe in the way Norman is stipulated to have done.[9] One would think that Norman ought to reason that, were he to possess such a power he would have seen evidence of it before, and that a lack of such evidence constitutes an undercutting defeater for the suddenly apparent belief.

2.3 The meta-reliability notion

So the non-undermining clause is the method Goldman adopts to handle the first two types of internal defeat. For the third type, simple defeat, Goldman takes an entirely different tack. Let us restate the problem: cases of negligent internal defeat are cases where some aspect of a cogniser's mental state eliminates their justification for a particular belief, but not because of the presence of positive or negative defeaters. These situations are always the result of some epistemic failure, past or present, on the part of the cogniser, but a failure that is not obviously reflected in the cogniser's mental state at the time of the target belief's formation. I gave an example, above, of an anterior failure with distrustful George. I will now provide an example of a concurrent failure, and then explain how Goldman handles such cases.

Imagine a superstitious lawyer who believes his client to be innocent, and is in possession of trustworthy evidence that conclusively proves this to be so. But the lawyer does not believe in his client's innocence because of this evidence, but rather he believes it because his fortune-teller tells him so, because at the time that the fortune-teller tells him so, the lawyer has a minor stroke, one that leaves him unusually susceptible to the persuasiveness of testimony.

I propose the superstitious lawyer case as an example of defeated justification,

specifically one of concurrent (or synchronic) simple defeat. Intuitively, the lawyer is not sensitive enough to the truth to be justified if he believes what he does for the reason stipulated. But note that the lawyer's belief in his client's innocence is not eliminated by positive or negative internal defeating circumstances. The lawyer in the example possesses no counter-evidence against his belief in his client's innocence, and he has plenty of trustworthy evidence in its favour. He has simply has some epistemic failure that causes him to believe in the fortune-teller.[10]

Concurrent simple defeat cases are almost usually cases where a cogniser arrives at a belief in some manner deemed inappropriate, when a better method was available (in some sufficiently attenuated sense) to the cogniser at the time. This category, combined with the diachronic epistemic failures exemplified by George, form the two sub-species of simple negligent defeat to be countenanced.

Goldman produces a number of different methods of dealing with the problem of simple defeat. In its earliest form Goldman's reliabilism was designed to handle cases of diachronic negligent defeat. On this early account, adumbrated in "A Causal Theory of Knowing",[11] a belief was justified just in case ". . . the fact that p is causally connected in the 'appropriate' way with S's believing P"[12] where appropriate is defined so as to exclude cases of anterior epistemic failure. This feature of historical reliabilism that gave it its name is successful for both internal (negligent) and external (non-negligent) anterior defeat.

Later on, in 'What is Justified Belief?'[13] Goldman modifies his earlier theory by restricting the analysis of appropriate causation to cognitive processes, preserving the

historical aspect, but removing his ability to appeal to historical mental events in the causal chain of a belief as grounds for its defeat. This change from historical-reliabilism to process-reliabilism initially forces Goldman to add a requirement that whatever cognitive process the cogniser uses to come to a belief, there must be no other more reliable process in the cogniser's possession such that, had she used it in addition to, or instead of, the original, she would not have come to the belief that she did. This stipulation is specifically designed to handle cases of negligent defeat, but it is only really successful at excluding cases of synchronic negligent defeat. Consider, what process(es) does George have available to him that, had he used it instead of or in conjunction with the ones he used, would have lead to him believing differently than he did?

In E&C Goldman revises his theory further to account for both types of negligent defeat by means of one general condition.[15] On the later account, Goldman handles these cases by enhancing the method of cognitive performance evaluation, and he does this by multiplying the number of cognitive objects germane to the evaluative process. Goldman introduces two other types cognitive processes, methods and second-order processes, that must also be evaluated when considering the justification of a given belief.

Methods are 'recipes' or 'algorithms' for the employment of basic psychological processes (such as memory and simple inference) to produce beliefs. Such heuristics are learned, not 'wired in' as it were, to the cogniser's mental apparatus, as the basic psychological processes are. Second order processes are, like their first-order cousins, basic psychological processes, built in and (largely) unreflective, but their task is specifically to acquire, evaluate and apply methods.[16] Goldman does not explicitly detail

the standards of measurement for these new entities, but he does propose that truth-attainment, suitably modified to the types of things being measured, is the correct one. And he offers suggestions as to what this meta-reliability requirement might mean for both of these new entities.

For example, Goldman proposes that a second-order process might be meta-reliable to the extent that it tends to increase the stock of reliable methods. Alternately, meta-reliability for a second order-process might mean that it must perform so as to increase the overall number of true beliefs of a cogniser, which might be accomplished by selecting few but very useful and reliable methods. Or meta-reliability might be cashed out to mean that a reliable second-order process only selects or retains methods that have a certain level of (first-order) reliability.[17]

As for methods, Goldman proposes that they be meta-reliable only if (1) they were produced and retained by reliable second-order processes, and (2) they themselves meet some standard of meta-reliability, such as only employing reliable basic psychological properties, or being reliably sensitive to certain epistemically relevant conditions, or being the most reliable method available.

This development allows Goldman to effectively handle both diachronic and synchronic cases of simple defeat. His reply to Putnam's example, for instance, shows both the flexibility of the new formulation, and its intuitive appeal. Putnam's acolyte is unjustified, on this analysis, either because the method he used in selecting a guide on truth and morals was unreliable, or because the second-order processes he used to garner that method were not meta-reliable, depending on how the story of the acolyte coming to

believe what he did is fleshed out.[18] This corresponds neatly to our initial judgement that Putnam's protagonist is unjustified by dint of his poor cognitive performance in arriving at his beliefs, by pointing out, in some general way, what went wrong, and why such an error is fatal to justification. Specifically, because the cogniser uses an unreliable method of belief formation, or learning. This analysis can also handle the problems of George and the superstitious lawyer. Both can be said to have unreliable second-order processes that are causally related to the target belief, and the lawyer obviously employs an unreliable method.

2.4 The problems with non-undermining and meta-reliability

Let us now turn to the problems engendered by these two methods, the non-undermining requirement and the meta-reliability requirement, that Goldman employs to handle the problem of defeated justification. To begin with 'undermining' as a concept is unexplanatory. Goldman does not provide us with a principle of undermining. He spells out three different types of circumstances that constitute undermining, but he does not give us an explanation of why they are such. Such a principle would be the lion's share of an explanation why justification is defeated in the types of instances cited. As it stands Goldman's inclusion of the non-undermining clause, while it might be technically successful, is itself insufficiently justified.

Further, the only plausible principle that might explain the inclusion of a non-undermining clause is that of the necessity of a rationality constraint on justification. But such a move, to impose a necessary condition of rationality on the justification of a belief, leaves the door open to all the problems that beset theories whose main justificatory

condition are rationalistic, such as the regress of justification. This is not a fatal criticism, but it does point out that a theory that could account for internal defeat without opening itself up to such problems would be well ahead of one, like Goldman's, which cannot.

Further, the underlying principle of undermining can only be, as I said, that of the necessity of a constraint of rationality for justification. But such a principle seems to require that we drop reliability as the sole criterion of justification. In essence, by the adopting of such a principle, Goldman tacitly adds the goal of epistemic coherence to the set of aims that underwrite his conception of knowledge and justification. While this new goal is not exactly incompatible with the original goals espoused by the reliabilists, namely the achievement of truth and avoidance of error, it is not of a piece with them, as it is not aimed at the truth, but rather at some level of belief compatibility. But the original motive for reliabilism, as for many of its truth-oriented cousins, was to eschew the traditional formulation of justification as a matter of evidence. With the inclusion of this clause, Goldman's later theory becomes in essence a hybrid, containing both external truth-oriented criteria and the more traditional rationality criteria. Again, this is not necessarily a fatal flaw. Indeed some theories unabashedly employ both types of criteria[19] and the truth of the matter might well lay down this path. But again, a thoroughgoing truth-oriented theory, one that employs only externalist principles, would be more internally consistent and less *ad hoc*.

It might seem initially possible for Goldman to eschew the non-undermining clause altogether and to rely on the meta-reliability requirement to handle all cases of internal defeat. Both positive and negative defeat cases can be handled by appeals to either the

unreliability of the method employed, or the unreliability of the second-order process responsible for the employed method. On this analysis, positive and negative defeat cases would become subsets of simple negligence cases, which the meta-reliability condition was designed to handle. They would be special types of simple cases wherein the epistemic failure results in a specific type of fault in the noetic structure of the cogniser at the time of the target belief's formation. So, for example, in the case of Samantha, we would say that she is unjustified because the second order process that garnered or vetted the method she used in arriving at her belief is unreliable, or because another, more reliable method was available to her, or, alternately, because the method itself was not sufficiently reliable, where this method is cast roughly as 'forming beliefs while disregarding rebutting defeaters'.

While this move does relieve Goldman of the problems with the non-undermining clause, it is not a very attractive alternative, because the problems engendered by the meta-reliability requirement are of a more severe nature than those plaguing the notion of undermining. To see this, re-consider the case of Samantha. If we wish to say that Samantha is unjustified because the method (and not the BFP-type) she used was unreliable, then we are confronted with the same problems I mentioned earlier, namely the parsing problem and the problem of psychological reality. As with BFP-types, methods and second-order processes are candidates for statistical measurement, and as such there must be types of them, and these types must be framed at a suitable level of generality. Further, again like BFP-types, individual method and second-order process tokens can fall under a variety of types, and we must judge which ones are relevant to the measure of

epistemic justification. Simply put, what the multiplying of objects of cognitive measurement does is multiply the number of problematic terms central to the theory, with each new object bringing with it its own baggage.

Of a piece with this undesirable result of the meta-reliability requirement is the worsening of the problem of psychological reality. In order to answer the charge of internal defeat, Goldman posits new and varied psychological objects, namely methods and second-order processes, whose statistical measurements will help determine the justificatory status of a belief. But this compounds the original problem of imposing on the realm of the special and general sciences by causing him to be even more dogmatically specific as to the nature of the mind. I will examine the problems of generality and psychological reality at length in the next chapter, but suffice it to say here that the meta-reliability condition makes process-reliabilism significantly more vulnerable to them.

We should also note that going the other way, and ridding ourselves of the non-undermining condition, is not successful either, since not all cases of negligent defeat can be accommodated by the non-undermining condition. As we saw earlier, the type of case we called defeat by simple defeat are notable by their absence of positive or negative undermining conditions, at least as Goldman outlines it. The existence of simple defeat cases mandates the use of another method. So PR is unsuccessful in handling the problem of defeated justification.

Chapter Three

Process Reliabilism and the problems with Belief Forming Processes

3.1 The parsing problem

The second type of charge that tells against process-reliabilism is that of the problematic nature of the notion of a belief forming process. In this chapter I examine ways in which problems with this notion arise, and I consider the reliabilists responses. The general charge is that the notions of belief-forming-processes and belief-forming-process-types, notions that do the epistemic work in PR, do not parse the world into processes and non-processes adequately for the purposes of answering questions about justification. I call this "the parsing problem". For our purposes Feldman's 1985 paper "Reliability and Justification"[1] can serve as the model of a successful method of arguing this point.

In the essay Feldman points out that what is lacking for process reliabilism to be a satisfactory theory of justification is a more detailed account of belief-forming-processes (BFPs) and their types. According to Goldman, a belief-forming-process is ". . . a *functional operation* or procedure, i.e., something that generates a *mapping* from certain states - 'inputs' - into other states - 'outputs' [. . .] [o]n this interpretation, a process is a *type* as opposed to a *token*."[2] For the sake of clarity, let us adopt Feldman's vocabulary and call a belief forming process token, an instantiation of a functional operation, a belief-forming-process (BFP), and the type a belief-forming-process-type (BFP-type). As

Goldman notes in the passage following the quotation, BFP-types are the types of things that can have statistical properties, such as being 80% reliable at producing true beliefs. It is the statistical measure of the truth-attainment ratio of the class of BFPs that a certain belief's actual causal process belongs to that determines that belief's reliability, and hence its justificatory status, on PR.

But, argues Feldman, a process can simultaneously be a token of many different process-types. By way of example, Feldman offers the specific BFP token leading to the belief that it is a sunny day. Such a token falls under all of the following different process-types: the perceptual processes, the visual processes, processes that occur on Wednesday, processes that lead to true beliefs, and on and on.[3] Presumably, argues Feldman, only one of these process-types are relevant to the justificatory status of the belief, yet all of them subsume the BFP-token in question. So what is needed is an account of how to go about determining the epistemically relevant process type, the one whose reliability measurement will determine the justification-status of beliefs.

But before we can even begin the search for the epistemically relevant process-types, continues Feldman, we must face a more basic problem that arises in parsing BFPs into types at all. This basic problem in parsing BFP-types is a dilemma between parsing them too broadly and too narrowly. Feldman calls the two horns the 'Single-Case problem' and the 'No Distinction problem'[4]. The former arises when process-types are defined too narrowly; the latter is the result of process-types that are too broadly defined. A process-type that is too narrow, one with a long and detailed list of criteria for its member BFPs, will have very few or a unique member BFP(s), thus disqualifying it as a

candidate for useful statistical measurement, the 'Single Case' problem. A process-type that is too broad, one that has perhaps only one or two criteria, will have too many member BFPs, and as a result will give rise to beliefs of obviously different epistemic status. But we would be unable to differentiate these beliefs by means of appealing to their justificatory status, because on PR they would share a common status by dint of their shared process-type origins, the 'No Distinction' problem. The problem of parsing BFPs into types such that they avoid this dilemma Feldman calls the 'Problem of Generality'[5].

3.2 Some possible solutions the parsing problem

Feldman considers a number of ways reliabilists might resolve this latter problem. Initially, he considers Goldman's attempt to parse BFPs into types by appealing to our common sense judgements. Feldman quotes Goldman, in "What is Justified Belief", saying: ". . . our ordinary thoughts about process-types slices them broadly"[5], and proffering: ". . . confused reasoning, wishful thinking, reliance on emotional attachment, mere hunch or guesswork, and hasty generalisation. . ."[6] as examples of process-types. Feldman terms this method of defining process-types the 'Standard View'. One might object here that the BFP-types that Feldman cites Goldman as proposing would not produce beliefs of interestingly different degrees of justification, since all of the beliefs issuing from their member BFPs would be unjustified. But consider hasty generalisation. Perhaps a person who comes to a very weak conclusion on the basis of very few evidentiary instances can be justified in her belief. Regardless, the classes Goldman proposes are useful by providing an example of the breadth of the categories that he suggests BFPs might be parsed into.

The problem that Feldman points out with this Standard View is that it engenders the No Distinction problem, because the process-types suggested are too broad. Feldman uses a set of examples where beliefs formed on the basis of vision is the process-type in question to demonstrate this. Consider (says Feldman) two different cases wherein a person comes to a belief that he sees a mountain goat. In the first case the person gets a good long look at the creature, from close up and in good light. In the second case the observation conditions are much poorer. The person is far away from the animal, the light is dim, etc. If we take vision to be the relevant process-type, as Feldman assumes the Standard View would, then we can't differentiate between the beliefs in the two cases by appealing to their different justificatory statuses, since they are both members of the same BFP category, and hence must have the same justificatory status. Yet they obviously have different degrees of justification. All else being equal, the mountain-goat belief is well justified in the first case and much less so in the second. Thus the proposed relevant process-type here is too broad. 'Vision' is too general a process-type to account for justification on a PR theory.[7]

One method that Goldman might adopt (Feldman continues) to narrow down process-types is that of relativising such types to external conditions, that is, facts about the world apart from psychological facts about the believer. So, for example, the mountain-goat belief in the first case would be the result of a different process-type than in the second one. In the former, the process-type would be something like 'forming vision beliefs under optimal conditions', while the latter would be an example of the process-type 'forming vision beliefs under sub-optimal conditions'.

This, as Feldman points out, would initially run directly counter to Goldman's proviso that the processes under investigation be cognitive processes that are internal to the organism. But, as Feldman also points out, this difficulty can be surmounted simply by making such external conditions factors that affect the reliability of process-types, rather than factors that help to determine the process-types themselves. In this way these factors would play a role in the measurement of a process-type's reliability, not in its definition, but the results for justification judgements would be the same as if they played the former role. The mountain-goat/vision cases, for example, would be handled by saying that the reliability of the process-type that gives rise to both beliefs varies with the 'epistemic conditions' attending the respective beliefs. In the first case, the 'vision' process-type has a reliability of, say, 90%, and thus the belief is justified, in the second case, the process-type has a substandard reliability rating of, say, 30%, and as a result, the belief is not justified.

Feldman's objection to this move is that it doesn't avoid the No Distinction problem, in the following manner. Any number of beliefs might arise as a result of a process-type that has an acceptable reliability rating in the given circumstances, and those beliefs may well have varying epistemic status, but again, on PR we would not be able to differentiate them on this basis. To take Feldman's example, if we return to the Mountain-goat scenario, in the first instance, the cogniser might form the belief that he was seeing an animal, as well as the belief that he was seeing a mountain goat. The former belief is more justified than the latter, since it is entailed by the second belief and it is a weaker claim, yet to the Goldmanian epistemic-condition-relative version of PR they are epistemically equal,

having both come from the same process-type under the same perceptual conditions.[8]

A possible counter to this objection concerning the No Distinction problem for an external-circumstance-relativised PR might be to further relativise the reliability ratings of process-types not just to external conditions, but also to types of beliefs. Feldman proposes that Goldman might answer the above charge by saying that the belief that 'that is an animal' is of a different type than that of 'that is a mountain goat', thus allowing us to assign them different justificatory status. But, Feldman argues, even with proper criteria for belief typification, which is wanting, the No Distinction problem still arises. Beliefs that are unarguably of the same type, arising under the same observation conditions, could still differ in justificatory status. Cases where a person is called upon to make judgments about the same type of thing, repetitively, like an umpire calling balls and strikes, to use Feldman's example, admit of varying of degrees of justification on the part of the beliefs formed. An umpire can certainly misjudge a pitch as well as get one right, and this might be for a number of reasons. The umpire might be biased, he might be tired, or inattentive, to name just a few possibilities. As a result of these conditions, the judgements he makes about various pitches might well differ in justification, but they are beliefs of the same type made under identical observation conditions.

What should also be noted is that both these moves, the relativising of reliability-ratings of BFP-types to external circumstances or to belief types, or both, also fail to avoid the other horn of the dilemma, the Single Case problem. If anything, such relativisation exacerbates this problem by making more things relevant to the justification of a belief, thus increasing the specificity of the reliability and justification determination based on

them. There are an infinite number of true facts about the circumstances surrounding any belief's formation, and if we include too many of them as relevant to its justification, call them 'epistemically relevant' factors, then we would have a reliability rating for a BFP-type that would be unique to the combination of that type with those circumstances, since those circumstances would be unique. If, in the mountain-goat cases, we were to describe in detail the observation conditions, down to the time of day, the exact position of the sun and other light-sources, the position of the observer, his height and age, etc., and include these things as 'epistemically relevant' circumstances, then the calculation of the reliability rating of vision as a BFP-type under these circumstances would be unique, and not useful in determining the same BFP-type's reliability under any other circumstances. The taking into account of external circumstances in the measurement of a BFP-type's reliability could in this way result in a different calculation of the reliability of a BFP-type for every belief issued by its member BFPs, and as such the purpose of having BFP-types at all, as categories for making useful statistical measurements, is defeated.

Consider further Feldman's umpire counter-example. The thrust of it is that the umpire makes judgements of different justificatory statuses, but the belief-type and epistemic conditions are held fixed. PRists might counter that the reasons listed for the umpire's misjudging a pitch, namely bias, fatigue and inattention, are examples either of different epistemic conditions or of different belief-types. For example, in the case of the fatigued umpire the PRist might say that, to the extent that the fatigue impairs the umpire's judgement, it is an epistemically relevant condition, and the reliability of the process-type in question must be assessed with this in mind. Similarly in the case of the

biased umpire, the PRist might argue that a belief issuing from a bias is of a different type than one issuing from purely perceptual data, and as such the reliability-rating of the BFP-type might well be different vis-a-vis the two types.

But both of these moves highlight the slippery slope that awaits the PR theorists who attempts to avoid the No Distinction problem by relativising reliability to internal and/or external circumstances. If we include factors such as the umpire's state of mind, his physical state, his mental effort, the lighting conditions and various and sundry other facts attending the beliefs formation, then the actual calculation of the reliability of the putative BFP-type in those circumstances will be a calculation of a situation so specific as to be unique. Not only would this defeat the purpose of parsing BFPs into types at all, since they would no longer be important factors in the determination of reliability and justification; but it would also be tantamount to collapsing justification to truth, because only true beliefs would emerge as justified on such exhaustive and precise calculations. On an exhaustively circumstance-relative PR, any circumstance, internal or external, that might result in a belief's being false, will be considered in the calculation of the BFP-type's reliability, and so only true beliefs would stand a chance of being justified.

3.3 Some more possible solutions to the parsing problem

In E&C Goldman proposes what he calls a ". . .promising lead toward a solution . .."[9] to the parsing problem. The suggestion is that what determines a belief's process-type is the narrowest description of such a type 'causally active' in producing the belief. Goldman asks us to envision a 'matching template' standard for vision beliefs as an example. A template is some pre-formed visual pattern that the vision process applies to

various inputs, and when the inputs match the template pattern to a sufficient degree, a vision belief is formed. His proposal is that the vision belief process-type actually active in a given situation is determined by the percentage of template-match present at the time of the belief. And that the reliability of these different vision processes is directly related to the degree of template match. So, if the person has a 90% template match, then the process-type in question is that of 'forming vision beliefs at 90% match', and the belief is then, say, 90% reliable, and therefore (probably) justified. We should note that this approach differs from the earlier approach, considered and rejected by Feldman, of relativising process-types to external factors, because the external factors here are not physically external, they are consonant with Goldman's desire to talk only of processes internal to the believer, since they are facts about the believer's actual mental processes, such as the degree to which her vision-template has matched with input, and the approach certainly seems compatible with a naturalistic theory.

The problem with this approach is that the idea of a 'template match' is not easily extendable to non-sensory beliefs. Goldman claims we could easily devise similar scenarios for other, non-sensory (what he calls belief-dependent) types of process-types. To do this we would have to conjure up some template-match-equivalent concept for each non-sensory type of process to determine what process-types are causally operative in a given instance. But while sensory beliefs are at least plausibly produced in a manner something like this model, where some set number of outlines shape most of the beliefs generated, it is not at all obvious that inferential beliefs are produced from anything like a manageable number of templates.

Another possible way of dealing with the parsing problem that Feldman notes in the literature is Schmitt's[10] proposal to reject the notion of BFP-types altogether, and define BFPs as simply the finite, physical sequence of events occurring in the believer that result in the formation of beliefs. The determination of reliability, on this account, would be done entirely counter-factually, based on estimates of a BFP's likely truth-ratio, were it to occur a number of times. This is called the propensity approach to reliabilistic justification, because in it justification is based upon the propensity of a uniquely appearing BFP to arrive at true beliefs in imagined situations. This move allows the reliabilist to solve the problem of generality by defining processes explicitly, so as to avoid the No Distinction problem, and, by appealing to the propensity of such strictly defined BFPs, to avoid the single case problem.

But, as Feldman argues, Schmitt's proposal engenders a variation on the single case problem in the following way: Because of the extremely explicit way in which Schmitt defines BFPs, a BFP can only produce the belief it does. And this means that any series of "completely specific events" in a believer that results in a true belief is a series that always leads to a true belief, and as such is a reliable process, and confers justification upon its one issuing belief. So if being hit on the head makes a person believe some truth of mathematics, in such a way that being struck this way would always result in such a belief, then that process is reliable, and thus the resulting belief is justified.[11]

We should also note that this objection can just as easily be raised against BFP-types as against discrete BFPs. Given what we know about the brain, it is not inconceivable, or even implausible, that a true belief might be caused reliably in a believer

by the direct manipulation of her brain. A person might come to reliably remember some incident in her childhood, for example, every time a rod was inserted into a part of her brain.[12] On what grounds would a PR theorists disallow beliefs arising under such circumstances status as a members of a specific BFP-type? If they cannot, then they must bite the bullet and accept such oddly formed beliefs as justified so long as the physical sequences produces true beliefs a sufficient amount of the time.

3.4 Another possible solution

Charles Wallis, in his 1994 paper "Truth-Ratios, Processes, Tasks and Knowledge"[13] takes up the defense of process reliabilism by proposing to relativise process-types to what he calls "specific cognitive tasks". Such tasks are delineations of cognitive performance requirements, bundled, on Wallis's account, into discrete units by cognitive scientists. Wallis recognises two problems for process-reliabilism: the problem of the proper characterisation of the process-types to be measured, and the problem of specifying the relevance class, or the class of situations in which the selected process-types are to be measured.

Wallis argues for the conceptual separation of the relevance class problem from the problem of generality, claiming that the former is properly a matter of determining what aspect of a cogniser's mental apparatus ought to be measured in a given circumstance, and that the latter is a matter of determining how it should be measured. Nothing in this discussion turns on this point for now, so for the duration of this exegesis I will continue with Wallis's formulation of the problems. Wallis's proposal is that both the characterisation of the process-type and that of the relevance class necessary to determine

its reliability ought to be determined by the cognitive task that the target belief falls under. The gist of this idea is that there are specific cognitive tasks that make up the mental functioning of a creature, and that these tasks are properly delineated by scientific investigation of the functioning of the creature in their normal environment. Wallis proposes three criteria to aid us in delineating these tasks: "(1) an idealised target function between input and output types, (2) a specification of the nomic correlations (including statistical correlations) that underlie the behavior of both the systems and the relevant objects within the domain, and (3) a specification of the relevant process by reference to a system's dispositions, viewing these dispositions as a strategy or set of strategies for generating outputs and other assorted behavioral responses (if any) from inputs from relying on certain nomic correlations."[14]

For Wallis, process-type charcterisation is performed by selecting the cogniser's behavioral (broadly construed to include mental behavior) dispositions across a task specification, and the reference class is determined by "the nomic correlations that underlies the system's performance of the task". In plainer language, Wallis proposes that we delineate tasks on the basis of a functional picture of goals and strategies to meet those goals, and then determine the reliability of a belief in a given instance by determining how well the strategies the cogniser has for performing the task that the belief's formation actually achieve the completion of the task, and that we determine this by testing these strategies (counterfactually) in cases where the actual physical relationships in the world that subtend the cogniser's abilities to complete the task are altered.

There are, of course, many questions to be raised here about the nature of this

task-discrimination procedure, and further questions might be raised as to the acceptability of the functionalist nature of the project itself, but leaving these aside, I believe that Wallis' approach is a step in the right direction. The only acceptable way to parse mental processes, for justificatory or any other purposes, is scientifically. The project of determining what aspects of mentation are active in belief formation, and how they do their jobs, is a project for the sciences, since it is a project of investigation of the physical world. Of course the goals of such investigations might be set by prior conceptual criteria, such that the delineation of the justificatorally relevant aspects of mentation might be targeted towards the eventual coincidence with our normative concept, but this role is formal, not pragmatic. It is the job of theorists to propose frameworks, wherein the ontological composition of the technical elements is left to the scientists. Wallis makes a move in this direction, but he retains, and even sophisticates the role of a scientifically suspect technical term, BFP-type, in his theory of justification.

Wallis proposes that the objects of justificatory measurements are 'strategies and sets of strategies' for solving problems. This is similar in intent to Goldman's move of introducing methods and second-order processes that are responsible for methods, as elements of justificatory measurement. Like Goldman, Wallis here desires to expand and clarify the aspects of a cogniser's mental apparatus that need to be measured for an accurate reliability/justification rating. According to Wallis, not only should we take into account the actual process that produced the belief, but also the 'nearby' processes that the cogniser also has to handle such belief production, the 'strategies' that the cogniser has for arriving at this type of belief. But, as we noted earlier, this method of

sophistication, while it is intuitively appealing, is fraught with problems. Wallis's theory does exactly what Goldman's does, namely, commit its proponents to an even more detailed and story of how the mind (or the brain, were theorists bold enough to couch there proposals at that level) works. A story about the division of mental labour into processes, methods (sets of strategies) and second-order-processes (wired-in strategies for obtaining, evaluating and retaining strategies).

My argument is not that this is an incorrect division of mental labour, that would be far too strong. Rather I contend that the formulation of justification in terms of the measurements of such detailed posited psychological entities is premature. It might well turn out to be the case that the mind works in the exact way, or in a sufficiently similar way, to that proposed by the process-reliabilists, but it might also well not be the case that this is true. I will not provide a detailed alternative method of mentation-parsing that makes process-reliabilism untenable, since I think that such a method can be easily arrived at, and I leave it as a project for the reader. The point is, even when process-reliabilism is pegged to scientific investigation, the positing of mental processes as real objects whose measurement determines the justificatory status of a belief is too strong a formulation. What is needed is a more general way of relating justification to cognitive performance, one that does not commit its protagonists to such a detailed theory of cognition.

I too want to enlarge the scope of reliability measurement to account for cases of negligent defeat, and to better accord with our judgements as to what is relevant to the reliability-measurement. However, the change I propose is to couch the theory at a more appropriately general level. In the following chapter I propose a theoretical framework

that explicates the concept of justification, without pinning us down to a possibly false picture of the mind.

Chapter Four

System Reliabilism

4.1 A review of the problems with Process Reliabilism

In this chapter I will briefly rehearse the arguments of the previous two chapters and summarise the problems with PR. I will then propose a solution to these problems, in the form of a new theory, system reliabilism, or SR. I will detail the theory, show how it differs from process-reliabilism, and explain how the concept of a test-set is to be fleshed out. Then, in the final chapter, I will argue that system-reliabilism can account for the types of cases, namely cases of negligent defeat, that plague Goldman's earlier version of process reliabilism. And that it can do so without having to posit theoretical constructs such as BFPs and the like, constructs that reliability theorists have had grave difficulty in adequately defining. First, let us review the problems with process-reliabilism detailed in the last two chapters.

Process-reliabilism suffers from two types of problems: 1) in Goldman's earlier, more basic, formulation the theory is unable to account for cases of negligent defeat (the problem of defeated justification), and 2) its proponents seem unable to clearly define the notions that are crucial to the theory's formulation, specifically BFPs and BFP-types. These two problems are related in that the attempt, on the part of the process-reliabilists, to amend the theory to answer the first charge results in a worsening of the problems that constitute the second one.

Process-reliabilists have only two options open to them when confronted with the

problem of negligent defeat. They can either simply require that beliefs not be negligently defeated in order to be justified, or they can elaborate the measurement of cognitive reliability so as to eliminate justification of internally defeated beliefs solely on the grounds of unreliability. Goldman makes both moves in E&C, by way of the non-undermining clause and the meta-reliability requirement, respectively. As I argued in the second chapter, the non-undermining clause is not in the spirit of a reliabilist theory of justification, as it is not truth-oriented, in fact it is completely *ad hoc*.[1] But in addition to this shortcoming, as a method of accounting for negligently defeated beliefs, at least as it is formulated by Goldman, the non-undermining condition is incapable of blocking the two sorts of simple defeat outlined in the second chapter.

Such cases mandate Goldman's second amendment of the theory, the meta-reliability requirement. But this change is more problematic than the first, as it worsens the second type of problem, that of the unacceptable nature of the psychological notions at the heart of the theory. It does so by multiplying the number of such notions threefold. In addition to BFPs and their types, Goldman introduces methods and second order processes as types of mental entities, and he requires that their reliability (or meta-reliability) also be measured in order for a determination of justification to be made. But this obviously compounds the problems inherent in using such notions, namely the parsing problem and the problem of psychological reality. How shall we define and parse methods and second-order processes? Surely they will be as difficult to define as BFPs, since surely a single belief can plausibly have come from an enormous number of methods, and a method from an enormous number of second-order processes, in the same way that a BFP

might be subsumed by any number of BFP-types. And of course these new entities force proponents of PR to espouse an even more detailed theory of the mind than previously. The theory now posits two new 'psychological mechanisms', that have very prescribed functional roles. Again, while cognitive psychologists and other mind researchers may well come to agree that this description of the structure of mentation is correct, with second-order processes garnering and vetting methods for solving problems, etc., to assume this before their results are in is premature.

4.2 System Reliabilism, a new direction

Let us try to diagnose what is wrong with process-reliabilism. To start with, it seemed that the basic, unsophisticated version like that found in Goldman's 'A Causal Theory of Knowing'[2], was unable to account for negligent defeat. Negligent defeat, again, is the defeat of justification due to some aspect of the epistemic comportment or mental state of the cogniser. The reason that this early version of PR is unable to account for this 'subjective irrationality', as Bonjour calls it, is because it focused too closely on the reliability of the issuing BFP-type, and disregarded the performance of other aspects of the cogniser's belief corpus and of their prior epistemic history.

The most interesting way that Goldman attempts to solve this problem is by elaborating the measurement of reliability to include other aspects of the cogniser's mental equipment. This broadening does seem to effectively solve the problem of simple defeat,

but at the expense of introducing two new notions central to the theory: methods and second order processes; notions that share all the definitional problems of the original BFPs and BFP-types.

One of these problems, the parsing problem, is the most damaging to PR. The reason PR cannot solve the parsing problem is that neither Goldman nor any other reliabilist offers an explanation of the notion of epistemic relevancy, a notion that is key to the parsing problem. As Feldman's argument show us, the question of what shall we put in our BFPs is really the question of 'what shall we take into account when considering whether a belief is justified', which is simply another way of asking 'what is relevant, *epistemically* relevant, to a beliefs justification?'

Consider how *ad hoc* Goldman's stipulation that only the cognitive processes of a person are epistemically relevant to the justification of that person's beliefs seems. Only the slightest reason is offered, namely that justification is a matter of "how a cogniser deals with his environment"[3], and as such evaluations of this sort should include only aspects of the situation that are internal to the cogniser. But Goldman proffers no general notion of epistemic relevancy. This lack is made more glaring when we see how PRists must retreat from this extreme position in the face of Bonjour's and Feldman's arguments, and find some way to include non-mental elements of a belief's formation, elements that intuitively seem epistemically relevant to the determination of justification, into the justification equation. But again, without an explanation of epistemic relevancy these moves are *ad hoc*. All this being said, reliabilism as a theoretical framework still holds many attractions for the naturalistically minded epistemologist. There are several key ideas in PR

that are correct, the first of which is a truth-oriented conception of justification. As well, PR conceives of justification as a matter of cognitive performance. A person is justified in some belief, on PR, just in case the BFP that they employed in the formation of the belief arrives at true beliefs some acceptable percentage of the time. This conception of justification as a matter of capacity, where the capacity measured is the attainment of truth, is the most promising linkage between truth and justification yet offered. The question is then; can we formulate a theory that uses this conception without falling into the traps that stop PR from succeeding?

I believe that we can. What I propose is to measure reliability (and therefore justification) in terms of the truth-attainment ratio of a cognitive system as a whole, and not that of the putative components of a cogniser's mental apparatus. This new theory, System Reliabilism, retains a truth-oriented, capacity view of justification, while at the same time obviating the need for a detailed theory of the mind. It does so by correlating reliability to the truth-production of the system as a whole in a variety of circumstances, actual or counter-factual, rather than to the truth-attainment ratio of whatever subset of the mind (or brain) is causally responsible for producing/sustaining the target belief.

System Reliabilism, then, is this idea: what it is for a belief to be justified is for it to have been formed by a cognitive system that is reliable enough at truth-attainment in a variety of epistemically relevant belief-scenarios. These scenarios form what I call the test-set of the target belief.

A belief-scenario is a description of the circumstances attending the formation or retention of a belief on the part of a cogniser. The examples and counter-examples that are

a mainstay of epistemological methodology can serve as paradigms. Bonjour's clairvoyants, Feldman's goat observer,[4] George the disbelieving student, all of these are belief-scenarios. A test-set is a set of belief-scenarios, actual or counter-factual, in which a person's truth-attainment performance is measured in order to determine her reliability, and hence her justification, for a certain belief. The test-set is just that, a test, albeit an idealised theoretical test, not one that must actually be passed to achieve justification. Specifically, it is a test of the believer's capacity to achieve the truth in conditions epistemically relevant to the belief under consideration.

4.3 The notion of a test and the principle of epistemic relevance

To a great extent, SR rides on the back of our understanding of the concept of a test, so perhaps we should elaborate on this concept. A test is just a measurement of capacity. We often measure a person's knowledge via a test, and this indicates to me that a notion of the justification condition of knowledge that makes it out to be the possession of a capacity or an ability on the part of the cogniser, an ability demonstrated by the passing of a certain test, is a promising one. Perhaps this would be best demonstrated by means of a detailed example:

Fred is a neophyte sonar operator. He begins his two weeks of training with an introduction to the machine, of which he was previously wholly ignorant. In the days that follow, Fred is taught the types of blips that different types of things make on a sonar screen. He is taught the visual and auditory cues that indicate surface vessels, fish, mountains, submarines, etc., as well as how to differentiate them from one another, and from background noise. He is also taught the range of possible interfering circumstances,

such as inclement weather, enemy jamming, ghost signals and the like. At the end of the course, there is a comprehensive examination, which Fred passes with flying colours. This test certifies his capacity to make sound judgments about objects in the water from the evidence of a sonar screen.

What Fred shows us that the idea of judging justification via a test is *prima facie* appealing. We commonly use tests to determine a person's capacity to make accurate judgements. And this is exactly what SR proposes to do. But what we need for SR is a method of answering the questions of belief-scenario definition and test-set selection: What composes a belief-scenario, why they are so composed, which ones are in the test-set of a particular one, and why? And these questions, like the questions about BFP definition and parsing that stymied PR, require a principle of epistemic relevancy. In answering these questions Fred's example can be of further use. If we consider Fred's certification exam as analogous to the test-set of a given sonar belief of his, then the questions on it are the equivalent of the individual belief-scenarios that compose that belief's test-set. So by examining how we go about setting the questions on Fred's test, what kind of questions we ask, and why, we might find a way to answer our questions about epistemic relevancy.

So what sort of questions would we expect to find on Fred's final examination? Well, the questions on this test are used to certify Fred's ability to make sonar judgements. As such, we would certainly expect some of the questions to be of the 'what is this blip?' sort, where Fred was asked to identify a given blip on a given screen in given conditions. This is an obvious test of the capacity sought in Fred, namely the ability to make these

judgements accurately. This is what the test certifies, that Fred is able to make accurate judgements about sonar blips, so testing him on this ability is obviously mandated.

As well, we might expect that some of the questions of this sort be 'trick' questions, that is, questions of the 'what is this blip?' sort that have a simple obvious answer that is incorrect because of some other aspect of the scenario. Here we are testing for Fred's sensitivity to the relevant conditions surrounding his judgements. We want to ascertain by these questions whether Fred could be relied upon not to make certain common errors in his judgements, errors like making a reading beyond the scope of his instrument, or under conditions in which his instruments are unreliable.

We might also require Fred to answer some basic questions about the nature of sonar. We might want him to know how sonar waves operate, and why they operate the way they do. As well we might expect him to demonstrate proficiency at controlling and operating the sonar machine, and perhaps even some minor diagnostic and trouble-shooting skills. This is so we know that Fred is capable of handling certain types of situations that might render his judgements unreliable.

And what do these questions all have in common? They all test Fred for capacities that are necessary for him to make accurate sonar judgements. For us to say that Fred knows that a blip on his screen is an enemy vessel he must posses certain abilities, and these abilities we test for on certification examinations. But why are the particular abilities tested for selected in any given case? Because having such capacities is necessary for successful (i.e. reliable) performance in the world. Given what we know about sonar, sonar machines, people and other aspects of the world and its workings, we have reasons

to think that these particular capacities underwrite Fred's capacity to make sonar judgements.

And here is the path that I think might profitably be followed. What we can adduce from our pre-theoretic judgements about Fred's test questions is that we already employ a rough and ready criterion for judging epistemic relevancy, and one that is completely naturalistic. To wit: something is epistemically relevant to a belief's justification just in case this thing impacts on those capacities that underwrite the person's ability to make truthful judgements, and we assess the measure of such an impact via our scientific (and other types of) knowledge. The types of questions that we find on Fred's test exemplify this criterion. We ask Fred questions that determine whether or not he possesses certain powers that are necessary for him to make reliable sonar judgements. Given what we know about the world, Fred is less (or more) likely to be reliable in a given sonar judgement depending on his possession of these capacities, and thus they are relevant in considering his justification for a given belief.

Let us try to translate this idea into a principle of epistemic relevance for SR. There are two places in the theory that require such a principle, in the formulation of belief-scenarios and in the selecting of belief-scenarios to compose test-sets. Again, the principle I propose is that epistemic relevance is a matter of the impact of a given thing on truth-attainment. So in determining what aspects of the world count towards a belief-scenario, we ask ourselves this question when considering whether some set of facts ought to be included: do these facts impact on the potential truth-attainment ratio of a cognitive system? Would the person involved be less or more likely to make a true judgement given

the presence or absence of the conditions listed in these facts, given what we know about the world?

To return to Fred, the presence of schools of fish that might be mistaken for ships we judge epistemically relevant, and therefore part of the belief-scenario (and part of a test-question) because Fred's truth-ratio might be affected by this fact. Whereas a fact about what Fred had for breakfast before making the particular sonar judgement at hand we judge not relevant, because it would not affect Fred's truth-attainment ratio. We would thus not list this circumstance as part of our belief-scenario.

And as for test-set selection, the principle works in a similar manner. Belief-scenarios are selected to the test-set of a particular belief on the basis of their epistemic relevancy to that belief. A scenario is epistemically relevant to a belief if, given what we know about the world, the person's performance in that scenario would demonstrate some capacity (or lack thereof) that impacts on the person's potential for truth-attainment in the target belief. So, for example, a question that describes a belief-scenario wherein Fred must make a judgement in bad weather we accept as relevant (and include on the test) because we feel that Fred's performance in this scenario is indicative of an ability that impacts his truth-attainment ability for normal target beliefs, namely his sensitivity to certain potentially defeating conditions.

One might object here that Fred's example is unfair, in that it deals with a particular, circumscribed bit of technical knowledge, one that is easily tested for. In order to see if SR works with the principle of epistemic relevance I have here laid out, perhaps we should examine a more mundane example of justification determination.

Let us take as our example Jane, a person who comes to the belief that something in her visual field is blue. In order to determine the justificatory status of Jane's belief on my account, we must compose a test-set for her belief, and in order to do this, we must select belief-scenarios that are epistemically relevant to that belief. This means that we would select scenarios that we believed tested for capacities that might impact on Jane's truth-attainment ability in the target belief. For example, we would certainly select scenarios in which Jane would make other beliefs about the colour of things in her visual field. This would be tantamount to asking Jane questions like "What colour is this?" while indicating a variety of objects. As well we would also have among our test-set scenarios ones where Jane comes to have beliefs on the basis of her vision. This would be very much like asking Jane to pass an eye test. And finally, we would include in our test-set scenarios in which epistemically relevant circumstances were present, circumstances that either change the way in which Jane must come to her colour beliefs in order for them to be accurate, or that remove any chance of her making an accurate (or accurate enough) belief to be justified in it. The former would be things like asking Jane to make colour judgements in different light, while the latter would be things like asking her to make similar judgements after putting on coloured lenses, or taking a hallucinogen, or something of the like. Again these scenarios are part of the test-set because they are epistemically relevant in the manner outlined. They test for a capacity that impacts on Jane's truth-attainment ability.

If Jane were to pass our test, that is to say if she could identify coloured things correctly often enough, and if her eyesight was in working order, and if she performed

well enough in making correct judgements in light of potentially defeating circumstances, either by withholding belief or by altering the manner in which she comes to believe so that she achieves a sufficient truth-attainment ratio, if Jane gets a high enough mark on her test, then we can say that she is justified, and therefore that, if her belief is true, she knows the thing she indicated to be blue.

This is by no means a definitive and exhaustive list of the belief-scenarios that might appear in Jane's actual test-set for her blue-belief. We might wish to test Jane for other capacities that are judged to underwrite her capacity to make color judgements, but I think the example has sufficient force the way it is. Surely, if Jane could pass such a comprehensive and exhaustive test, she would be a reliable judge of colour, and so justified in her individual colour beliefs.

4.4 A modification to the principle of epistemic relevance

As I noted earlier, scenarios that embody defeating conditions for a belief are on this principle of relevance, epistemically relevant to that belief. But perhaps not universally. Again, defeating conditions are conditions that, if present, remove justification for a given belief from a cogniser. They fall into two broad categories: negligent and non-negligent. I discussed negligent defeating circumstances above, in chapter two. Non-negligent defeating circumstances are circumstances where knowledge, and therefore justification, on my account, since we hold the circumstances fixed with regard to the truth and belief requirements of knowledge, is vitiated by something that lies outside the cogniser's epistemic responsibility. Gettier cases are the paradigm. These are cases wherein a cogniser validly deduces a true proposition from justified premises, but where

our intuitions tell us that the belief in question is not a piece of knowledge for that cogniser. Let us briefly describe such an instance.

I will use one of Gettier's original examples for this purpose.[5] A man has two friends, Smith and Jones. The man believes that Jones owns a Ford. He has good reasons for holding this belief. He has seen Jones in a Ford many times, and Jones has told him that he (Jones) owns a ford, and so on. From this belief, the man deduces that the proposition "Either Jones owns a Ford, or Smith is in Barcelona" is true, since it is a disjunct in which one of the disjuncts, namely "Jones owns a Ford", is true. The man has no idea where his friend Smith is, and he simply plucks the phrase "Smith is in Barcelona" out of the air. As luck would have it, Jones does not, in fact, own a Ford. However Smith, by coincidence, is in Barcelona at the time the cogniser's belief is formed. Thus the disjunctive proposition expressed by the cogniser's belief is true, since one of its disjuncts is true, but we feel strongly that the cogniser does not know it to be true. In this case it seems that the justification for the cogniser's belief has been vitiated, but not by any aspect of the cogniser's behaviour, nor by his mental state. Rather some other element of the situation, something beyond the cogniser's ken, has acted so as to vitiate the justification, and therefore the knowledge, of the belief for the cogniser.

The bizarre nature of non-negligent defeating circumstance examples points to a potential problem with our formulation of the principle of epistemic relevancy. Consider another scenario wherein a believer's mind is removed by aliens, altered so that the believer would consistently make a mistake about a type of judgment, and replaced in the believer's head, all without their cognizance, some time prior to the believer's formation of

a belief of that type. While these circumstances would certainly impact on the truth-attainment ability of the person, they seem too far-fetched, too unlikely to occur, and too removed from the cogniser's ability to take appropriate measures if they were to be the case, to be considered epistemically relevant to the target belief.

It would seem, then, that not all types of belief-scenarios embodying defeating conditions for a belief are epistemically relevant to that belief. The scenarios might well embody circumstances that are far too attenuated to bear on the justification of the belief at hand, even if they would defeat the belief's justification were they to occur. So we should modify the principle of epistemic relevance of belief-scenarios to one another slightly, to say that scenarios should be selected for a test-set on the basis of their impact on the truth-attainment ability of the believer for the target belief *given that the believer ought to be sensitive to the defeating circumstances present in the potential scenario*. This, at the very least, this means that the test-set should not include scenarios wherein the defeating circumstances are extremely unlikely. There is no point in testing people for capacities they will never need.

4.5 Some of the benefits of System Reliabilism

I will now briefly examine some of the benefits of this type of test-set framework theory. To reiterate: For a system-reliabilist, a belief is justified just in case the system that produced the belief is reliable at forming beliefs of that type, and this reliability is measured in terms of the system's performance in the test-set of the belief, such that a system is reliable if it attains some ratio of true-to-false beliefs in the 'normal' scenarios composing the test set, and withholds belief often enough in the defeating-circumstances

scenarios.

The first thing we should note is that the theory is completely naturalistic. The determination of justification is made entirely based on our knowledge about the workings of the world. Unlike Goldman's earlier PR, SR does not require an *ad hoc* non-undermining clause, and unlike some versions of that theory it does not artificially limit what is pertinent to justification to poorly understood, hidden mental processes. SR is also external all the way down. It does not require that in order for a belief of hers to be justified, a cogniser must have 'privileged access' to what justifies it. As such it avoids the problems inherent in internalist theories. Again, this is in contrast with earlier, non-undermining versions of PR.

As well, SR is bolstered by our common practice of testing for knowledge. Exams, certifications, auditions, all of these are demonstrations of the fact that we use a rough and ready version of SR in many of our day-to-day judgements about justification and knowledge. Consider a case of asking a young child the time. If you were unfamiliar with the child, and did not know whether he could tell time or was simply making up a number (in other words, doubting the justification the child has for his belief), the way to go about ascertaining whether the child is justified would be to put him through a little test. To Ask him questions such as 'where are the hands of the clock?' or 'where did you see the time?'

As well there are a number of pleasant outcomes stemming from the use of the test-set notion, as I have defined it. One is that a theory formulated along its lines can be flexible with regard to the actual constituent elements of the test-set. While I argued for naturalistic, capacity-oriented criteria above, the framework can accommodate other

criteria for test-set selection. The types of criteria advanced will of course reflect the proposer's views on what they think is epistemically relevant for testing when we make justification judgments, but the framework itself stipulates only that some set of scenarios composes the test-set, and that a certain performance in that set is required for justification.

A test-set theory could also be neutral with regard to the manner in which test-set scenario-selection-criteria are determined. The test-set notion allows for the use of scientific and societal directives, and so a theory formulated along its lines can accommodate all these aspects of knowledge, and reflect them in its formulation, as it should.

Yet another advantage of this framework is that it allows us to tailor our tests to the system in question. This allows us to accommodate our intuitions about the variability and relativity of justification, as well as the possibility of knowledge being held by less than fully functional adult cognisers. System Reliabilism is consistent with the belief that the actual standards of knowledge may vary from person to person, and from humans to another species. This malleability stems once again from the framework's neutrality vis-a-vis the actual make-up of the test-set. It allows for different test-sets, given different motives and assumptions about the system.

For these reasons, I believe that the test-set notion is an important and useful tool in forming justificatory theories. Regardless of the correctness of SR as a theory, the notion of the test-set and its role in determining justification is one that I think has promise. In the next and final chapter, I will examine how SR, by the use of the test-set

notion, handles the problems that plague PR, and I will further explain the benefits of the structure of SR.

Chapter Five

System Reliabilism, solutions and problems

5.1 System Reliabilism and defeated justification

In this chapter I will examine how this new theory, system reliability or SR, handles the problems that stymied the older theory, Goldman's process reliability, or PR. I will then consider some of the possible objections to the new theory, and give some arguments in its defense. I will also re-visit and expand on some general arguments for the adoption of an externalistic, naturalistic stance on justification, and for reliabilism and SR in particular.

I believe SR will allow us to respond to both of the two broad classes of problems that proved fatal to the older theory: the problem of defeated justification, with its sub-problems of positive, negative, simple and non-negligent defeat, and the problems with the concept of a BFP, including the problem of psychological reality and the parsing problem. Let us consider the problem of defeated justification first.

As we recall, cases of negligent defeat are cases wherein a cogniser holds a belief, a belief that may have come to them in a reliable manner, the justification for which is vitiated by some aspect of the circumstances that the cogniser is epistemically responsible for. Goldman's earlier version of PR was unable to account for negligent defeat, since it made no mention of the cogniser's mental resources beyond those causally active in the production or sustainment of the belief.

In his later theory, Goldman amends his formulation to include a 'non-undermining' clause, specifically to address this problem. The non-undermining clause is tantamount to an admission of a necessarily internal component to justification on Goldman's part, but, as he formulates it, it is incapable of handling cases of what I have called simple defeat, cases like that of George in chapter two.

What makes cases like George's difficult for the earlier Goldman's PR is that George does not have internal defeaters, in the form of rebutting, undercutting, or an absence of supporting beliefs, so George's belief is not 'undermined', on Goldman's explication of the term, by his mental state. This failure on the part of the non-undermining condition to handle cases like these are what prompts Goldman to propose the sophistication of the reliability measurement, with all of its attendant problems. So PR strikes out in the realm of negligent defeat. But how would SR handle such cases?

An SR theorists would say this: As scenarios that are defeating circumstances, cases of positive or negative defeat would certainly be scenarios that would test the sensitivity of the system to the truth of the type of belief in question. As such, they would qualify under the epistemic relevancy criteria as potential test-set members.[1] And in test-set cases where defeating conditions that a cogniser ought to be sensitive to obtain, any behaviour but the cogniser's withholding of belief will inevitably lead to failure. Consider Fred once more: a defeating condition question on Fred's certification test would be one that stipulated a scenario, say a blip at too far a range, or in an impossible location, where the only correct way to answer would be to disavow knowledge of what the blip represented. If the question was phrased in the form "What would you say if the Captain

asked you what this blip represented under these circumstances ?" it would require Fred to answer something like "I can't tell, sir" in order to be correct.

This is what an agent ought to do in a defeating circumstance scenario. What we test for when we include these scenarios in our set is the ability of the agent to doubt when doubt is called for. The reason for this is the same as that of the original principle: We believe that the world operates a certain way, and we further believe that we have at least some grasp of the way in which the world works. This knowledge, which spans the gamut from scientific investigation to the individual experiences that ground induction, allows us to predict the way the world will work in a variety of counterfactual circumstances. One such set of circumstances are defeating circumstance scenarios, and our knowledge about the way the world works tells us that, in some unacceptably high percent of such cases, if the cogniser were to form the type of judgement called for when such circumstances obtain, their judgements would be false.

It is this eventual failure of a cogniser who persists in making judgements when defeating circumstances are present that grounds our judgements of unreliability for agents like Bonjour's. A believer ought to doubt, when defeating circumstances are present, and she ought to do so precisely in order to maximise the probability of her belief's being true. In order to pass our test a cogniser must not make a judgement at all, or she must doubt any such belief that she forms unwittingly, in test-set scenarios that embody defeating circumstances, since she will get it wrong far too often if she does not.

So system-reliabilism's answer to the problem of negligent defeat is this: if the defeating conditions present in a given scenario are of a sort that would be exemplified by

other scenarios included in the test-set of the belief in question, and if the cogniser's behaviour in the scenario is indicative of some epistemic habit, so that they would perform similarly on some percentage of the test-set, then the cogniser will fail to achieve a high enough grade in the test-set to be reliable, and therefore justified, in the target belief.

Take Bonjour's cases again as examples of positive and negative internal defeat. If Norman and Maude's cases are ones that embody defeating conditions for the belief in question, then cases very much like them would be included in the test-set of the current belief, since these conditions test for capacities that are epistemically relevant to the judgement being made. If the agents were to arrive at beliefs of the sort described in the manner described in these cases as a matter of habit, then a system reliabilist can say that they are unjustified in their beliefs because they are insufficiently reliable.

If Maude's belief about the whereabouts of the President is positively internally defeated by a set of strongly held beliefs of hers, and yet she still stubbornly continues to hold it, and if this is something that happens to her all the time, then she will surely fail our test. There will certainly be a number of scenarios in our test-set that will embody circumstances like this one. And in these sorts of cases, in order to pass the test, Maude must withhold belief, since they indicate defeating circumstances that Maude ought to be sensitive to, in this case, circumstances of massive counter-evidence. If Maude were to habitually come to judgements in the face of such circumstances, then she would form false beliefs far too often to pass our test.

Similarly for Norman, who believes, as a result of his clairvoyant power, and without any evidence either for or against, the belief about the president's location that

suddenly pops into his head. If this is an example of Norman's standard epistemic behaviour, he will surely not pass our test. Again, we will select scenarios that will test for sensitivity to this type of defeating condition, lack of evidence, on the basis that such cases are epistemically relevant to the present case, since forming beliefs under these conditions is going to impinge on a cogniser's ability to make correct judgements. If Norman comes to a belief despite the presence of these circumstances then he would be making a judgement where none ought to be made, and he would be wrong an unacceptably high percentage of the time.

So how does system-reliabilism handle George, my example of anterior simple defeat? Well, we can see that if scenarios like that in question were to be considered members of the test-set for the belief, then we could answer the charge in the same manner which we employed for cases of positive and negative internal defeat above. So first we must ask ourselves, is this scenario a defeating condition for the belief in question? It would seem so, but in order for it to qualify as a test-set member on our criterion, it must also be a scenario that embodies a type of defeating condition that we feel the cogniser ought to be sensitive to. In this case, it certainly seems within George's ability to ascertain that his habit of eschewing introductions might be epistemically damaging. So it seems that cases like the one George finds himself in would be included as members of the test-set of this belief, and as a result, if we take George's failure as indicative of a common (to some degree) set of events, then George is not sensitive enough to these types of defeating conditions, and as a result, is unreliable and unjustified.

So for SR, the failure of these protagonists to perform correctly in certain test-set

scenarios demonstrates an insensitivity, on their parts, to the truth in those sorts of situations. And this lack reduces their reliability at making judgements of that sort sufficiently to remove their justification for the one in question.

We can glean two things from these arguments. One is that certain types of defeating circumstances might apply to a wide range of belief types, and that scenario's embodying these circumstances, call them radically defeating circumstances, might be found in the test-set of most, if not all beliefs. Scenarios involving strongly held negatively relevant beliefs to the target belief on the part of the cogniser, for example. It would seem that any belief should have scenarios that test for the cogniser's sensitivity to such conditions in its test-set.

The other thing to note is that where judgements differ as to the relevance of a candidate scenario, so will judgements of justification. If we feel that George need not be sensitive to the type of defeating condition outlined in the example, then our test-set will not contain similar scenarios, and consequently George might well pass our test, and be justified. Those that feel George ought to be sensitive to such circumstance might well fail him, judging him unjustified. The same reasoning applies, *mutatis mutandis*, to the stricken lawyer, our case of concurrent simple defeat.

5.2 Some more problems with defeated justification

There are two further issues to be dealt with before we finish with the problem of defeated justification. First, we can imagine a Bonjour-type scenario, where a clairvoyant comes to a belief in the face of defeating circumstances, but where the clairvoyant is more sophisticated and only comes to beliefs in this fashion when they are the result of her

clairvoyant power. We can further imagine that the clairvoyant is utterly reliable in determining when a belief of hers is a result of her power. In cases like these, it seems that our motive for marking her poorly on the part of our test that deals with such situations, situations wherein a belief of hers issuing from her power flies in the face of massive counter-evidence, has no purchase.

Remember, our reason for positing the failure of such agents was that, given the way the world works, an agent will be wrong far too often if she consistently makes judgements in the face of defeating circumstances, such as massive counter-evidence. But here in this modified Bonjour case our agent will (almost) always get it right in these sorts of scenarios, since she will only form such beliefs when they come to her as a result of her reliable clairvoyant power.

How shall we handle such sophisticated clairvoyants? To answer this question, we must go into further detail as to the noetic structure of our sophisticated clairvoyant. The problematic cases are ones wherein, unlike Bonjour's protagonists, our clairvoyant does not suffer from either positive or negative negligent defeat. This means that she has no defeating counter-evidence with regard to the accuracy of her power, and she has sufficient evidence that she has such a power. Because, of course, in cases where these circumstance do not obtain she will be unjustified in her beliefs for these reasons, regardless of the actual reliability of her power, in the same way that Bonjour's clairvoyants are.

But if this is the case, then there is no reason to deny the sophisticated clairvoyant knowledge. If the sophisticated clairvoyant has some evidence in her favour and no

damning evidence against her belief, and she is actually reliable in such beliefs, then she is justified. Such a clairvoyant simply has a new method of forming beliefs, and that the test-set for beliefs that are formed in this manner should reflect this, and thus should be comprised of belief-scenarios designed to test for those capacities and sensitivities needed for reliability in the target scenario. On this 'clairvoyant-belief' test-set, then, our cogniser would have to demonstrate both truth-attainment in her clairvoyant belief, as well as sensitivity to the defeating conditions that pertain to her power, if such exist.

The second concern is Gettier problems, or cases of non-negligent defeat, an example of which I gave in the preceding chapter. I do not claim here to have a definitive answer to this problem, but in that respect SR is no worse off than any other candidate theory of justification. This is an heroically difficult problem, but I think that Goldman sheds some light on what might be the correct response to it.

The root of the problem facing SR and other justification theories in Gettier examples is that we intuitively wish to avoid assigning knowledge to people who come to believe the truth 'accidentally', because this seems tantamount to admitting the existence of something like 'epistemic luck'. These are people, like the man in the Smith and Jones example, who hit upon the truth accidentally, even though they are reliable enough in the main to pass the test for the given belief scenario. Intuitively, they don't seem to *know* their beliefs, in spite of their reliability, because they hit upon the truth luckily. In his 1988 essay "Strong and Weak Justification"[2] Goldman points out that 'epistemic luck' might not be all that repellent a concept, and he offers in support of this the fact that 'moral luck' and 'legal luck' play similar roles in their normative arenas.[3] I take Goldman in this article

to be suggesting the following: in scenarios where an agent would pass our test of reliability, yet her justification is defeated through no fault of her own (non-negligent defeat, in other words), the agent might still be said to have *knowledge*, if she gets at the truth. Of course, we would want to keep this group to a very small percentage of all successful knowledge claims. On an SR account this would surely be the case, given that negligence is determined by performance in the test-set of the scenario, and a failure in this arena will result in a defeat of justification and knowledge, regardless of the epistemic luck attending the incident.

So SR's answer to the problem of defeated justification is this: if an agent ignores defeating circumstances as a matter of habit, then she will fail to make the grade on the test-set of a target belief-scenario. They will fail to achieve a high enough truth-ratio in scenarios embodying defeating circumstances for that belief because such defeating circumstances will too often result in an incorrect judgement. For cases of non-negligent defeat, like Gettier cases, a failure in the test-set will still remove justification. But for a small number of cases, where the agents pass the test and get to the truth, they owe their resulting knowledge to 'epistemic luck'. This in the same way that a reckless driver who avoids injuring himself or others owes his freedom (of conscious, if not of person) to his 'moral luck'.

5.3 System Reliabilism and the parsing problem

As for the second type of problem, that of defining the central concepts of PR, BFPs and BFP-types, there are two separate problems here, the problem of psychological reality and the parsing problem. Let us tackle the problem of psychological reality first, as

its treatment is much shorter than the parsing problem.

SR responds to this challenge by relieving us of the need to posit such entities at all. For SR, what makes a belief justified is the performance of the overall cognitive system that produced it, and not the performance of that system's individual psychological components. Thus we need not worry about which parts of a cogniser's mental landscape ought to be measured. So much for that problem.

The parsing problem, however, will not be put to rest as easily. It would seem that SR inherits its own version of the parsing problem. The argument would go like this: By testing for a cogniser's capacity to arrive at true beliefs in a test-set reliably, we have simply transformed all of the problems surrounding the definition of BFP-types to those of test-sets. According to SR, justification judgements are arrived at via a measurement of the cogniser's performance in achieving true beliefs in a certain test-set, but, the argument goes, these sets cannot be parsed so as to avoid the no distinction/single case dilemma. Just as Goldman's PR couldn't parse BFP-types, such that they were neither two broad nor too narrow, so will SR be unable to formulate test-sets that avoid the same problems. The problem would be that the test-sets of beliefs we arrive at will either be too broad, admitting too many varied beliefs-scenarios to accord with our intuitions, or they will be too narrow, so as to contain too few elements to be epistemically useful.

As I said in the fourth chapter, this problem is underlain by another problem, the problem of epistemic relevancy. Both Goldman and I attempt to parse things so that the resulting entities are useful for making justification judgements. Goldman selects BFPs, whereas I propose test-sets. In order to make such categories one must provide criteria for

the determination of epistemic relevancy between the members of the putative type or test-set. On both accounts, the classes of things are grouped together because they are epistemically relevant to one another. On Goldman's account, to say that a belief is produced by a certain BFP-type is to say that all the epistemically relevant features of the process that formed it were considered in the formation of the type. Similarly for my theory, to say that a belief is a member of the test-set of another is to say that it is epistemically relevant to it.

As we recall, in Goldman's theory there was no principle of epistemic relevance given, and the only plausible (non-arbitrary) method of solving the problem was by appealing to science. But for Goldman at least, I argue that this is not tenable, since the sciences he needs (neurology, psychology, and all the sciences in between these two extremes) have not yet reached a useful verdict in the areas he needs them to. What's more, they might never reach a verdict that he finds acceptable, since they might eventually produce a picture of the mind/brain that is incompatible with the functionalist construal mandated by PR.

SR, however, has an acceptable principle of epistemic relevancy, and it can appeal to it in solving the parsing problem. On SR, epistemic relevancy is a matter of what our scientific best-guesses tell us about the types of capacities an agent needs to make a judgement, and the impact of certain facts upon those capacities. The types of inquiries we might make in this pursuit are much more likely to be easily and correctly answered by scientists than questions regarding the functional nature and ontology of the mind or brain. That a certain amount of light is needed to make a colour judgement, for example, or that

sound waves are distorted by water in such-and-such a manner, these will be the types of contributions that I foresee science making in our determinations of epistemically relevant circumstances. And these are contributions that I think our scientific knowledge will have, if not exactly no trouble providing, then certainly less so than providing what Goldman must demand, namely information on the workings of the mind.

So SR answers the parsing problem by appealing to the principle of epistemic relevancy. We determine what scenarios are relevant to a particular belief by determining if their respective attributes impact on a cogniser's ability to make correct judgements. And we appeal to science to tell us what those capacities are, and what circumstances would affect them, given how they operate in the world.

It should be noted, however, that the principle of epistemic relevancy that SR invokes would not do all its work *via* scientific investigation alone. Our social practices and beliefs must play a role as well. For example, someone in our society might demand that a certain type of scenario be included on a test-set, say one that tests for the sensitivity of the agent to the possibility of mendacity on the part of testifiers, where someone from a more trusting culture would not deem such a scenario epistemically relevant.

I think that it is obvious that societal practices can and should play some role in a theory of justification, not the only role, not even necessarily the primary role, but some role. And this is as it should be, since justification is at heart a normative concept. It would be incorrect of any theory of justification to shut the door completely to social facts. Determining whether a person is justified in a belief is inescapably a human practice, done

almost exclusively on other humans, and quite often for socially dictated purposes. Thus we must allow facts about these things to influence our judgements. In SR our social beliefs and practices aid us in determining questions of epistemic relevancy, alongside scientific knowledge. The belief- types we make will thus be useful in arriving at justification judgements, because they will be produced by a complex interweaving of influences, one that mirrors the complexity of the task of parsing the world into human kinds.

5.4 A summary of arguments in favour of System Reliabilism

In this final section, I will outline the general merits of a reliabilist conception of justification, and specifically for one that employs a test-set notion, as I have defined it here.

Reliabilism is an externalist theory of justification. This means that it eschews the accessibility requirement on justification-conferring properties. Whatever it is that justifies a belief, for an externalist, need not be within the ken of the cogniser. Following Bergman's[4] framework, we can place the later Goldman in the category of moderate externalism, meaning that at least one of the justification conferring properties named by the theory is not necessarily accessible to the cogniser. Externalist theories gained popularity after the advent of Gettier's problem in part because they allowed theorists to answer Gettier's challenge by appealing to criteria outside the ken of the cogniser to explain why the protagonists in the examples lacked knowledge.

The traditional internal/external split between justificatory theories, made glaring by Gettier cases and other examples of what I have here called non-negligent defeat, runs

very deep, and I think it reflects a basic tension among theorists as to what justification means, as a functional term. Internalists, exemplified in theories of people like Chisolm[5] and Lehrer[6] are wont to believe that justification consists in the reasons a cogniser has for believing what they do. To be justified is to have good reasons to believe, and to have good reasons to believe is to have sufficient evidence of the belief's truth.

Externalists seem motivated by a different conception of justification. The theories of Armstrong[8], Dretske[9], and Goldman, to name a few, are attempts to make justification out to be a link between the two other conditions of knowledge, truth and belief. What Externalists want to say is that justification is a matter of a belief being held by a cogniser, because, in some way, the truth of the belief *caused* the cogniser to believe it. At least this is how many of the theories actually formulate their justification conditions.

There are, of course, many reasons other than this underlying intuition that Externalists might adopt such a stance. There are many grave problems with internalist theories, problems like the regress of justification, the problem of meta-knowledge, and others. These problems stem from internalism's restriction of justificatory conditions to those that cite other beliefs of the cogniser. This locks justification into the cogniser's mind, and as a result, it has no footing that cannot be undermined. There seem to be no basic beliefs of the sort that would satisfy the foundationalists attempt to solve these problems, so we are left either with circular justification models, coherentism, or externalism.

And internalist theories also seem untenable for explaining basic sensory beliefs. The justification for my belief that the bus is red, for example, intuitively does not seem to

rest with the other beliefs I have. It certainly doesn't seem to rest on beliefs about the appearance of things being red meaning that there is a good chance that they are red, and about the lighting conditions being conducive to proper colour judgement, etc. Basic sense-beliefs simply do not seem justified by any other beliefs, but rather by other factors, factors such as the accuracy of the sense modality, and the conditions of the perception.

It is no accident that externalist theories of justification often take as their case examples sensory beliefs. External theories, on the whole, handle sense-beliefs rather easily, because they can appeal directly to the external factors that seem epistemically relevant. But externalism has its share of problems. To begin with, earlier external theories, such as Armstrong's, and the early Goldman's, make the link between truth and belief that constitutes justification too strong. Armstrong's nomological model, for example, requires that there be some scientific law relating the belief in the cogniser's mind with its truth.[10] Goldman's historical reliabilism requires a proper causal history of the belief, where proper is spelled out so as to link the truth of the belief with the fact that the cogniser holds it.[11]

That the condition is too strong is demonstrated by the difficulty externalist theories have in handling negligent defeat cases. The conditions generated by many externalist theories are satisfied by cognisers of the Bonjour sort. We can assume, to appease Armstrong, that there is a law of the universe operating so as to give Maude true clairvoyant beliefs. We can stipulate that, if the president were not in New York, Maude wouldn't have believed it, *pace* Nozick.[12] And we have already seen how Goldman's theory fails to handle cases like Maude's.

So the need arises to modify the type of causal connection between the truth of a belief and its being held that would constitute justification so as to address these problems. There were a number of candidate emendations to the standard causal/nomological link proffered. The largest branch were probabilistic theories.[13] Probabilism, as it sounds, is the relating of a belief's justification to the probability of its truth. A variety of these theories was put forward, but it was Goldman's PR that won prominence as the most promising theory type.

Goldman's theory proposed a causal link that was statistical, rather than nomological. The reason a belief is justified, on Goldman's PR, is because its causal mechanism is statistically successful in arriving at the truth. This statistical model of the link allows PR to avoid the problem of counter-instances that plague stronger formulations, like all probabilistic models did. But Goldman's proposal is superior to other models, such as Bayes', in terms of explanatory power. Saying that justification simply amounted to the probability of the belief's being true did not speak at all to our many, firmly held intuitions about the concept at all. Reliabilism, on the other hand, accords with a number of these intuitions, as I outlined in the first and fourth chapters.

But Process Reliabilism is not successful in explaining cases of negligent defeat. Further, it forces us to assume a model of cognition as a precondition of the application of the theory. I propose that the answer to this problem is changing the objects of justification measurement. For a number of reasons, both theoretical and pragmatic, I think it better to measure the truth-attainment ability of a cognitive system as whole, rather than the putative mental components that produce beliefs within the cogniser. Such

a theory would allow us to take into account the whole of the cogniser's mind, rather than just a portion, and so allow us to account for negligent defeat. Pragmatically, the methods we have for measuring behaviour of whole cognitive systems far exceed the methods we have for brain/mind analysis, in both number and accuracy.

And finally, the test-set notion, central to SR, is my method of making this move. Justification remains a capacity tested for, as on Goldman's account. Indeed, it remains the same general capacity, truth-attainment. But the test is one administered to the cogniser, rather than to her mental constituents taken individually. This type of test is one that could, more than conceivably, be administered, as Fred's example illustrates.

I believe that this alteration to the framework of a reliabilist theory, this change from process to system, and the resulting change in test methods, will focus us in on the correct questions, at least, for the search for the explication of the justification concept. The question of epistemic relevancy, as I have here defined it, is central to such an explication. Thus even if SR, as I have formulated it, is substantively incorrect, I believe that, as a framework for future, the notion of the test-set will prove fruitful.

Endnotes

Notes to Chapter 1

1. I should begin by noting that this essay will only focus one type of knowledge, specifically propositional knowledge, or what is sometimes called 'knowledge that' as opposed to 'knowledge of' and 'knowledge how'. This is not to be taken as a disagreement with nor a dismissal of these aspects of the concept of knowledge.

2. The discussion of the proper definition of knowledge is to be found in the final part of the *Theaetetus*, secs 200-210.

3. Alvin Plantinga, in his book Warrant, the Current Debate (N.Y., Oxford University Press, 1993) makes this point in his interesting history of internalism in epistemology (see pages 11-15).

4. There are some intricacies that I delicately skip over here, concerning the exact formulation for this 'accessibility requirement' of internalist theories. These concerns are not very pertinent to my essay, but for those interested in these questions, William Alston, in his essay "Varieties of Privileged Access" in *American Philosophical Quarterly*, vol. 8, no. 3, 223-41 (1971) examines a number of different formulations of the internalist restriction in detail.

5. Again, I will not comment on the debate about how to properly categorise the precise epistemic status justifying beliefs must have. And again Alston investigates this issue in some depth, this time in his essay "Internalism and Externalism in Epistemology" in *Philosophical Topics*, vol. 14, no. 1, 118-196 (1986). He does agree however that the requirement I give here (in a rough and ready form) is the most plausible.

6. For a classic (if somewhat convoluted) example of foundationalism, see Roderick Chisolm's Theory of Knowledge , 3rd ed. (Englewood Cliffs, New Jersey, Prentice Hall Press, 1989).

7. For an example of a detailed coherence theory, as well as a primer on what is wrong with foundationalism, see Laurence Bonjour's The Structure of Empirical Knowledge (Cambridge Mass., Harvard University Press, 1985).

8. See John Pollock's 1986 book Contemporary Theories of Justification (New Jersey, Rowan and Littlefield, pubs.), pp 75-83 for an amplification of this and other arguments against coherence theories.

9. Edmund L. Gettier, "Is Justified True Belief Knowledge ?" in *Analysis* 23, 121-3 (1963). Gettier's problem, as it has famously come to be known, also provides many externalist theories with problems, including Reliabilism, the one which I endorse, as we

shall see.

10. D.M. Armstrong, Belief, Truth and Knowledge (Cambridge, Cambridge University Press, 1973).

11.Goldman credits the initial idea to F.P. Ramsey, in his book Foundations of Mathematics and Other Logical Essays (London, Routledge and Kegan Paul, 1931). As well, Goldman acknowledges that the term 'Reliabilism' was in the air before his first paper on the topic, "A Causal Theory of Knowing" appeared in *The Journal of Philosophy* 64, pp 771-791 in 1967. Most notably, D.M. Armstrong propounds a variant of reliabilism in his book Belief, Truth and Knowledge.

12 Alvin Goldman, Epistemology and Cognition (Cambridge, Mass., Harvard University Press, 1986).

13. Alvin Goldman, "What is Justified Belief" in Justification and Knowledge, George Pappas, ed., (Netherlands, Kluwer Academic Press, 1976), p. 116. The pagination references are from Goldman's collection of essays entitled Liaisons (Cambridge, MA, MIT Press, 1992).

14. Laurence Bonjour, "Externalist Theories of Empirical Knowledge" in *Midwest Studies in Philosophy*, vol 5, Studies in Epistemology, P. French, T. Uehling and H. Wettstein, eds., (Minneapolis, University of Minnesota Press, 1980) pp 53-74.

15. Ibid., p. 60.

16. Richard Feldman, "Reliability and Justification" in *The Monist*, Vol. 68, no. 2, 169-174 (1986).

17. See Plantinga's Warrant the Current Debate, p. 198.

Notes to chapter 2

1.Found in his "Externalist Theories of Empirical Knowledge", pp 59-62. The examples are here somewhat quickly presented for the purposes of this essay, but their crucial aspects are intact.

2. The vocabulary of defeaters, undercutters, overriders, nuetralisers was introduced and expanded upon by Alston, in various papers. See especially "Internalism and Externalism in Epistemology" in *Philosophical Topics*, 14, no.1 (1986) and "Concepts of Epistemic Justification" in *The Monist*, 68, no. 1 (1985).

3. Goldman, E&C, pp 110-1. It should be said that Goldman does not explicitly categorise defeaters as I am doing. What leads me to believe that he sees these examples as members

of another type of defeating circumstances is his method of responding to them, as I will discuss further later on in the chapter.

4. Ibid., p 63. In E&C Goldman introduces the idea of framing a justification theory in terms of Justification Rules (J-Rules, as he calls them). He has a number of arguments for this type of structure, to be found at the beginning of the fourth chapter (p 59). I don't go into this here, since it is slightly orthogonal to my interests.

5. Ibid., p110.

6. Ibid., p 111.

7. Ibid., p 112.

8. Ibid., p 113.

9.Ibid., p 112.

10. Although the settings and characters of this scenario are reminiscent of Keith Lehrer's 'Gypsy Lawyer' example, structurally it is a variant of the 'epistemically serendipitous brain lesion' examples that Plantinga provides in Warrant, the Current Debate p 192, 199, and I acknowledge the debt.

11. Alvin Goldman, "A Causal Theory of Knowing" in *The Journal of Philosophy* 64, 357-372 (1967) .

12. Ibid., p 80.

13. Alvin Goldman, "What is Justified Belief" in Justification and Knowledge, George Pappas, ed.(Netherlands, Kluwer Academic Press, D. Reidel, pub., 1979).

14. Ibid., p. 123.

15. Goldman, E&C, see pages 51-53.

16. Ibid., pp 93-95.

17. Ibid., pp 113-116.

18. Ibid., p 110.

19. For examples of these 'hybrid' theories, see William Alston's essay "An Internalist Externalism" in *Synthese*, 74, 265-83 (1988) or consult Marshal Swain's book Reasons and Knowledge (Ithica, N.Y., Cornell University Press, 1975).

Notes to Chapter Three

1. Richard Feldman, "Reliability and Justification" in *The Monist*, Vol. 68, no. 2, 169-174 (1986).

2. Goldman, "What is Justified Belief?", cited in Alvin Goldman, Liaisons, (Mass., M.I.T. Press, 1992) p 115.
3. Feldman, p. 160.

4. Ibid., p.161.

5. Goldman, "What is Justified Belief", cited in Feldman, p. 161.

6. Ibid., p. 162.

7. Feldman, p.166.

8. Ibid., p.164.

9. Goldman, E&C, pp 50-1.

10. Frederick Schmitt, "Justification as Reliable Indication or Reliable Process" in *Philosophical Studies*, 40, 109-17 (1981).

11. Feldman, p. 169.

12. The thought is not all that far-fetched. Consider Wilder Penfield's findings that direct stimulation of a patients brain could result in spontaneous memory recall. See his book Mysteries of the Mind (Princeton, NJ, Princeton University Press, 1975).

13 Charles Wallis, "Truth-ratios, Processes, Tasks and Knowledge" in *Synthese* 98, no. 2, 243-270 (1994).

14. Ibid., p. 265.

Notes to Chapter Four

1. I should note that I am not alone in leveling this complaint against Goldman. Bonjour "Externalist Theories of Empirical Knowledge" and Plantinga in Warrant, the Current Debate make the same arguments.

2. This is not a failure exclusive to PR. Most if not all of the early externalist theories could not handle the problem of defeated justification. Notable among these are Armstrong's and Dretske's.

3. Goldman, "What is Justified Belief", p. 116.

4. Actually, Goldman introduces the goat observer example in "What is Justified Belief?", and Feldman simply modifies it to suit his purposes.

5. Edmund L. Gettier, "Is Justified True Belief Knowledge ?" in *Analysis* 23, 121-3 (1963).

Notes to Chapter Five

1. Things are not always this clear, as we saw in our discussion of non-negligent defeat in chapter four.

2. Alvin Goldman, "Strong and Weak Justification", in *Philosophical Perspectives*, 2, *Epistemology*, 1988, J.E. Tomberlin, ed. (Atascadero, CA, Ridgeview Pub. Co., 1988).

3. Ibid., p. 139.

4. Michael Bergman "Internalism, Externalism and the Defeating Condition" in *Synthese* 110 (3) 399-417 (1997).

5. Roderick Chisolm, Theory of Knowledge, 3rd edition, (Englewood Cliffs, New Jersey, Prentice Hall Press, 1989).

6. Keith Lehrer, Knowledge (London, Oxford University Press, 1974).

7. D.M. Armstrong, Belief, Truth and Knowledge (Cambridge, Cambridge University Press, 1973).

8. Fred Dretske, Seeing and Knowing (London, Routledge, & Kegan Paul, 1969).

9. This is especially obvious in Armstrong and the early Goldman of "A Causal Theory of Knowing".

10. cf. Armstrong's Belief, Truth and Knowledge, p. 166: ". . . there must be a *law-like connection* between the state of affairs [S's believing that p] and the state of affairs that makes p true . . ."

11. cf. Goldman's "A Causal Theory of Knowing", p. 80: "S knows that p if and only if the fact that p is causally connected in the 'appropriate' way with S's believing that p."

12. Robert Nozick, Philosophical Explanations (Cambridge MA, Harvard University Press, 1981).

13. Plantinga Traces the influence from Thomas Bayes through such writers as F.P. Ramsey, Rudolph Carnap, Thomas Nagel and others, cf Warrant, the Current Debate, p.114, note 2.

Bibliography

Books:

D.M. Armstrong, Belief, Truth and Knowledge (Cambridge, Cambridge University Press, 1973).

Robert Audi, The Structure of Justification (Cambridge, Cambridge University Press, 1993)

Laurence Bonjour, The Structure of Empirical Knowledge (Cambridge Mass., Harvard University Press, 1985)

Roderick Chisolm, Theory of Knowledge (Englewood Cliffs, New Jersey, Prentice Hall Press, 1989 -3rd ed.)

Fred Dretske, Seeing and Knowing (London, Routledge, & Kegan Paul, 1969)

Alvin Goldman, Epistemology and Cognition (Cambridge, Mass., Harvard University Press, 1986)

Keith Lehrer, Knowledge (London, Oxford University Press, 1974)

Alvin Plantinga, Warrant, the Current Debate (N.Y., Oxford University Press, 1993)

John Pollock, Knowledge and Justification (Princeton, NJ, Princeton University Press, 1974)

John Pollock, Contemporary Theories of Justification (New Jersey, Rowan and Littlefield, pubs. 1986)

F.P. Ramsey, Foundations of Mathematics and Other Logical Essays (London, Routledge and Kegan Paul, 1931).

Marshall Swain, Reasons and Knowledge (Ithica, N.Y., Cornell University Press, 1975)

Robert Nozick, Philosophical Explanations (Cambridge MA, Harvard University Press, 1981)

Articles:

William Alston, "Varieties of Privileged Access" in *American Philosophical Quarterly*, vol. 8, no. 3, 223-41 (1971).

Willliam Alston "Internalism and Externalism in Epistemology" in *Philosophical Topics*, vol. 14, no. 1, 118-196 (1986).

William Alston, "Concepts of Epistemic Justification" in *The Monist*, 68, no. 1 (1985).

Laurence Bonjour, "Externalist Theories of Empirical Knowledge" in *Midwest Studies in Philosophy*, vol 5, Studies in Epistemology, P. French, T. Uehling and H. Wettstein, eds., (Minneapolis, University of Minnesota Press, 1980).

Edmund Gettier, "Is Justified True Belief Knowledge ?" in *Analysis* 23, 121-3 (1963).

Richard Feldman, "Reliability and Justification" in *The Monist*, Vol. 68, no. 2, 169-174 (1986).

Alvin Goldman, "A Causal Theory of Knowing" in *The Journal of Philosophy* 64, pp 771-791 (1967).

Alvin Goldman, "Discrimination and Perceptual Knowledge" in *The Journal of Philosophy* 73, 771-191 (1976).

Alvin Goldman, "What is Justified Belief" in Justification and Knowledge, George Pappas, ed., (Netherlands, D. Reidel, pub., 1979).

Alvin Goldman, "Strong and Weak Justification", in *Philosophical Perspectives*, 2, *Epistemology*, 1988, J.E. Tomberlin, ed. (Atascadero, CA, Ridgeview Pub. Co., 1988).

John Pollock, "A Plethora of Epistemological Theories" in Justification and Knowledge, George Pappas, ed., (Netherlands, D. Reidel, 1979).

Ernest Sosa, "The Raft and the Pyramid", in *Midwest Studies in Philosophy*, vol 5, Studies in Epistemology, P. French, T. Uehling and H. Wettstein, eds. (Minneapolis, University of Minnesota Press, 1980).

Frederick Schmitt, "Justification as Reliable Indication or Reliable Process" in

Philosophical Studies, 40, 109-17 (1981).

Charles Wallis, "Truth-ratios, Processes, Tasks and Knowledge" in *Synthese* 98, no. 2, 243-270 (1994).

Anthologies and Collections:

Philosophical Perspectives 2, *Epistemology*, 1988, J.E. Tomberlin, ed. (Atascadero, CA, Ridgeview Pub. Co., 1988)

Midwest Studies in Philosophy, vol 5, Studies in Epistemology, P. French, T. Uehling and H. Wettstein, eds., Minneapolis, University of Minnesota Press, 1980)

Justification and Knowledge, George Pappas, ed., (D. Reidel, pub., 1979)

William Alston, Epistemic Justification, (Ithica, NY, Cornell University Press, 1989)

Alvin Goldman, Liaisons, (Cambridge, MA, MIT Press, 1992)