

University of St Andrews



Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

Zen and the Fine Art of Forming True Beliefs
- The Philosophers Guide to the P's

Lars Bo Gundersen, 2000
Ph.D. Thesis
Department of Logic and Metaphysics
University of St. Andrews



Zen and the Fine Art of Forming True Beliefs
- The Philosophers Guide to the P's

by Lars Gundersen

Declarations

I, Lars Bo Gundersen, hereby certify that this thesis, which is approximately 74.000 words in lengths, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

7/6/00 __

I was admitted as a research student in June, 1995 as a candidate for the doctoral degree; the higher study for which this is a record was carried out in the University of St. Andrews between 1995 and 1999.

7/6/00 __

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the doctoral degree in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

7/6/00 __

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

7/6/00 __

Abstract

In Chapter I I argue that prephilosophical intuitions about knowledge and epistemic warrantability are best captured in modal terms. Epistemic skill is not merely a matter of getting things right in the actual world but, rather, a matter of getting things right in a relevant range of hypothetical scenarios. In this sense epistemic skill is reminiscent of the fine art of archery: skill in archery is not measured by actual instances of hitting the target but rather by a general capability to do so, also if circumstances had been modified slightly. Chapter II is an exposition and criticism of the two most influential theories within modal epistemology: those due to Fred Dretske and Robert Nozick. I examine their theoretical strength in the light of a number of classical difficulties within epistemology. Likewise I will survey the points of criticisms launched against these theories. In particular I consider criticism of the specification of epistemological methods; the counterfactual semantics underlying the theories; and, lastly, the various sort of counterexamples that, at least *prima facie*, appear to undermine the plausibility of these theories. I conclude that these difficulties are not necessarily ascribable to the core ideas underlying Dretske and Nozick's theories but rather are the result of their specific formulations—in particular their exact formulations in counterfactual terms. My core thought is that the theoretical edifices in Nozick and Dretske's theories, if comprehended in terms of dispositional rather than in terms of counterfactual reliance, remain their theoretical strength while avoiding various pitfalls. In particular, recent work on dispositional indicates that the taxonomy of counterexamples to the tracking theory precisely mirrors the taxonomy of counterexamples to the counterfactual analysis of dispositional properties generally. This indicates that the various sorts of classical counterexample to the tracking theory are precisely those which are to be expected when one, misguidedly, attempts to subject a—in nature—dispositional concept to counterfactual analysis. These findings in Chapter II are used for a more detailed discussion of dispositional in Chapter III. Drawing from results in recent philosophical projects which also trade on dispositional I am led to the idea that it is feasible to develop a significant amendment of modal theories of knowledge and warranted beliefs in dispositional terms. In Chapter IV I focus on the particular difficulties accruing from a conditional analysis of dispositional properties, the instances of 'masked', 'mimicked' and 'altered'

dispositions. It will be argued that some of these difficulties can be remedied once it is recognised that alleged counterexamples of this kind to the conditional analysis have to meet certain metaphysical and semantic constraints, notoriously the constraints I will define in the two principles of Variable Strictness and Compositionality Strictness. In Chapter V I will examine the possibility of a dispositional theory of knowledge in the light of these two principles. The driving thought here will be that knowledge is the result of exercising a dispositional skill to form veracious beliefs—just like competent archery displays as the result of exercising a dispositional skill to hit the target. This theory will lead us to a revised diagnosis of the problem cases for the classical modal theories and, eventually, to a more sophisticated version of a dispositional epistemology. In chapter V I will also survey some of the virtues inherent in such a dispositional theory: its theoretical strength in combating scepticism, its capability to clearly diagnose the classical problem cases for theories of knowledge and, not least, its upshot in an alternative way of viewing the structure underlying the wealth of beliefs whose aggregate constitutes our world-picture.

Acknowledgements

The list of people to whom I—one way or another—have come in debt during the process is far too long to be recorded here. I owe special thanks to Crispin Wright who has been an inexhaustible source of inspiration throughout, for his never failing encouragement, patience and support but, most of all: for being an example. To Andy Law for keeping me alive, sane and even reasonable fit, for huge quantities of valuable P's and, most of all: for his 'sight' training. To the Southgate Group for numerous stimulating discussions and everlasting friendships. To Matthew Herbert and Jakob Hohwy for their great friendships and support. To my family for their love and encouragement. And to Fraser McBride for invariably letting me win in backgammon when I needed it most. Finally to Marguerite Nesling for her splendid proof-reading.

The work was made feasible by a generous scholarship from the Danish Research Academy (excellently administered by Birgit Christensen and Tinna Meyer) and, in its initial stage, by a grant from the Sophus Berendsen Fund. It has been a very special privilege to carry it out at Department of Logic & Metaphysics in St. Andrews.

For my beautiful angel

Contents

Chapter I

The General Idea Underlying Modal Epistemology	1
---	---

Chapter II

Modal Theories of Knowledge and Warranted Belief	17
2.1 Penetrability of Epistemic Operators	19
2.2 Conclusive Reason	28
2.3 Four Misgivings	41
2.4 Tracking	47
2.5 Knowing More by Knowing Less	65
2.6 Summary	71

Chapter III

Tracking Dispositionality	73
3.1 Epistemic Dispositionality and Tracking	74
3.2 Dispositional Concepts	76
3.3 Altering, Masking and Mimicking	84
3.4 Epistemic Mimicking	87
3.5 Epistemic Masking	90
3.6 Epistemic Altering	94
3.7 Internal Versus External Dispositionality	100
3.8 Summary	102

Chapter IV

Masking, Mimicking and Altering	105
4.1 Mimicked Masking — Subjunctive Conditionals	107
4.2 Establishing a New World Order	114
4.3 Counter-mere-antecedent-factuais and Counter-entire-factuais	125
4.4 Mimicked Masking — Dispositional Claims	128
4.5 A Naive Solution	130
4.6 Lewis on Naive solutions	135
4.7 Mimicked Masking Contexts — Quantifying In	142
4.8 Altering Altering	147
4.9 Summary	158

Chapter V	
A Dispositional Theory of Knowledge and Warranted Belief	163
5.1 Review of the 'Problem Cases'	168
5.2 A Dispositional Theory of Knowledge and Warranted Beliefs	174
5.3 Tracker Conditionals versus Tracking Conditionals	181
5.4 Certainly Disposed	187
5.5 A Warrant for Tracking	194
5.6 Scandalous Scepticism	197
5.7 Transmission of Warrant	20
5.8 Aetiological Foundationalism	212
5.9 Summary	215
Bibliography	221

Chapter I

The General Idea Underlying Modal Epistemology

Once upon a time it seemed a simple matter to define knowledge: an agent S knew that P, the contention then was, if and only if S had a justified true belief that P. But then Gettier came along and reminded everyone that justified true beliefs do not always suffice for knowledge.¹ Gettier-type considerations can be illustrated by the following example: the Clock case. If Foucault, whose clock happened to stop exactly 24 hours ago, forms a belief as to what time it is, then, although this belief is justified (by looking at his clock) and true (the clock actually shows, just now, what time it is), it by no means counts as knowledge. Foucault is extremely lucky that he gets the time right in this instance. Had he consulted his stopped clock 5 minutes earlier, or later, he would have got the time wrong. And knowledge, properly understood, does not allow for such an amount of good fortune.

Gettier's reminder that knowledge cannot be identified with justified true belief kicked off a heated debate about what knowledge *really* is, and, further, whether knowledge allows for conceptual analysis at all. In the course of this debate it has been suggested that each of the ingredient conditions in the classical tripartite definition—belief, justification and truth—calls for revision, both in order for these ingredients jointly to suffice for knowledge, but also, and maybe more

¹Gettier [1963]. Gettier was not the first to publish counterexamples to the classical tripartite definition of knowledge. Lewis Carroll's *A Tangled Tale* contains suitable candidates (thanks to James Chase for pointing this out). Gettier was not even the first to publish Gettier counterexamples in a strictly philosophical context. The clock case appears in Russell's *Human Knowledge*, published in 1948 (p.113).

surprisingly, for each individually to count as necessary. The motivation driving the present inquiry is that some conceptually illuminating stories yet remain to be told. I will assume, as a starting point, that knowledge is related, somehow, to belief, justification² and truth and regard the challenge as to clarify and improve on these ingredients' exact nature and internal association.

In order to avoid the attack Gettier launched against the classical tripartite definition it seems reasonable to focus attention on modalised theories of knowledge. Whatever other requirements a theory of knowledge must meet, it has to account for the fact, hinted at above, that an accidentally true belief does not amount to knowledge. In order to know, one has to get things right in a manner that secures truth, not merely in the actual situation, but also in a range of hypothetical situations. For example, for Foucault to know what time it is, we would demand of Foucault that he got things right, even had things gone slightly differently and his clock had stopped 5 minutes later than it in fact did.

No doubt the most influential modalised theory of knowledge has been the so-called 'tracking account' developed by Dretske (1970) and (1971) and Nozick (1981). The basic idea behind this theory is that mere justification is not appropriate for gaining knowledge. The supporting justification has to be of a certain kind. In particular, it must be constituted in such a manner that were it to be constituted in an identical, or very similar manner in slightly modified scenarios, it would still hit the target, as it were. When we direct our beliefs towards worldly truths justification serves as a guide. The metaphor of belief as an arrow is helpful here. Every time an arrow is discharged, its direction and

²Historically, the term 'justification' has been associated with internalist accounts of justification, that by which the cognitive agent can justify his belief in P if challenged to do so. There are, however, no a priori grounds for thinking that any such justificatory feature of the epistemic situation has to be internal to the agent. The use of the term 'justification' is by no means meant to imply that this issue between epistemic internalism and externalism is foreclosed.

destination are determined by a complex and interacting system of steering devices; the initial angle of the arrow relative to the intended goal, the tightness of the bowstring, the design of the steering feathers, the weight of the arrow, etc. etc. And whether the arrow hits the intended target depends, to a large extent, on the constitution of this steering device. But not exclusively, for there are also certain device-external factors involved in determining the arrow's direction such as the force of the wind, the mobility and size of the target, etc. Now, in order to be a master of archery it is not sufficient merely to be capable of manipulating the steering device such that the arrow hits the target under a particular specifiable external condition. It does not even suffice to be capable of manipulating the steering device in such a manner that the arrow hits the target repeatedly as long as these specified circumstances are held constant. Even a blind person might accomplish that given he has been appropriately trained. However, the slightest change in device-external surroundings would result in such a blind person missing the target, no matter how otherwise skilled he is. A true master, therefore, is one who, in manipulating the steering device, takes such potential changes of external factors into account. His manipulation of the steering device is so constituted that, were it constituted identically, or very similarly, under varying circumstances, he would still hit the target. The manipulation of the steering device must be *sensitive* to the device-external factors. Likewise, the thought is, in modal theories of knowledge, the processes by which a belief is formed have to be so constituted that, were they to be identically, or very similar constituted, in slightly modified hypothetical scenarios, they would still lead to a true belief. A justification for the claim to know the time consisting in consulting a stopped clock does not satisfy this requirement. Granted, the evidence so obtained might lead to a true belief. It might even do so repeatedly given it always is constituted under these particular circumstances (that consultation occurs at a particular time of the day). But were Foucault to base his belief on this evidence in slightly different circumstances, for example circumstances in which the clock stopped 5

minutes earlier than it did, his belief would no longer be directed towards the truth.

This way of presenting matters raises the following question: what is it for belief forming procedures to be constituted in a certain way and, *a fortiori*, what it is for justification to be identically, or very similarly constituted in various hypothetical situations? And it is in attempting to answer this question that Dretske's and Nozick's theories ultimately reveal their deficiency. Since, without resources to answer this question a modal epistemology loses its explanatory power, we take it to be one of the major tasks of our investigation to search for a suitable response.

One of the issues that this question raises concerns internal vs. external standards for justification. Someone in favour of internalism will argue that every relevant feature of the constitution of an agent's justification must be accessible to that agent's cognitive apparatus—must be, as it were, open to view for the agent. In the clock case an internalist would thus claim that the features relevant to a discussion of what constitutes Foucault's justification are such features as it appearing to Foucault that there is an apparently well-functioning clock in front of him and, furthermore, that the perceptual conditions seem, to Foucault, to be well suited to investigating what time it is by consulting the clock. An externalist, on the other hand, would claim that these internal constituents to Foucault's justification, although relevant, have to be augmented by factors exceeding the scope of what is immediately open to view for Foucault. An externalist would thus insist that Foucault's justification must be constituted also by the facts that there really *is* a clock in front of him, that it *is* well functioning (if it is) and that perceptual conditions for consulting the clock really are appropriate—all this independently of how matters might or might not appear to Foucault.

Here again the analogy between belief and an arrow is helpful. I said that the archer's manipulation of the steering device has to take into account potential changes of device-external circumstances. The manipulation must be constituted in such a way that it is sensitive to these device-external factors. The archer must

somehow, consciously or unconsciously, read these device-external factors—wind, speed of the target, etc.—and calculate how the steering device should be calibrated accordingly. Competent archery can be considered a two-way system. It involves both a passive intake of external clues and a controlled (outgoing) dispatch of the arrow. The archer receives clues about factors external to the steering device, manipulates the steering device accordingly, and becomes, thereby, capable of sending the arrow towards its target. The archer can be said to hold a repertoire of default adjustments suited for various standard situations—default adjustments which he is continually ready to make to meet relevant external factors.

Is there any analogous structure to be found in the case of belief formation? Well, in order to answer that question it is important to clarify what the analogous components are. I suggested the arrow may be compared with belief and the steering device with justification. But what is the counterpart of the factors external to the steering-device? What is the counterpart of the circumstantial variables, modification of which influences whether a particular adjustment of the steering mechanism guides the arrow to the target? In order to answer these questions we need to determine what the circumstantial variables are, the modification of which influences whether a given piece of justification issues in a true belief. And those that naturally come to mind are the agent's epistemological settings. If we accept the internalist account of justification we can say that a piece of evidence might guide an agent to the truth in one setting, and maybe even say that it might always guide the agent towards truth in that setting. But we can still claim that, were this setting to be slightly modified, that piece of evidence would no longer be trustworthy. For example Foucault, since he has recently consulted his clock, is (internally) justified in believing what time it is. And if he is put in a setting which has it that clocks are well functioning, that piece of (internalistic) justification will invariably (or almost invariably) lead him to the truth. But if he were to be put in a different setting which has it that clocks are not generally

reliable—were he for instance in a museum for old clocks—then that particular piece of evidence might not be any good for him in sorting out what the time is. And if Foucault misreads his own settings, if he is not sensitive to clues about which settings he happens to be in, the evidence might even become manifestly misleading.³

But, you might object, surely Foucault cannot be mistaken about his own settings. It is platitudinous that he is aware of what settings he happens to be in. Of course he knows whether he happens to be at home perceiving his own clock or in a museum for old clocks. Likewise he would be extraordinarily inattentive if he hadn't noticed whether his clock is at least moderately reliable. And more generally it seems that his beliefs regarding his other settings—that he is not asleep etc.—‘second-order’ beliefs as we might phrase them, will always be appropriate, and thus he always will be capable of calibrating his (internalistic) justification relative to the right settings.

This is at least how we usually conceive of these matters. As in archery, it makes sense to consider our epistemic relation to the world as a two-way system. We are

³It is no easy matter to state, in any formal manner, what settings are. The core idea underlying the notion is that almost invariably there is a gulf between a given piece of (internalistic) evidence (as e.g. when it appears to Foucault, from consulting his clock, that it is 6 o'clock) and the state of affairs supported by this piece of evidence (e.g. it is 6 o' clock). We could, if we wished to, state a very long list of collateral states of affairs which have to obtain in order for that piece of evidence to be veridical. (e.g. Foucault is not sleeping, conditions are suitable for perceptual investigation, the clock is well functioning ...). If the entire list holds true, circumstances are conducive to the investigation in question. This fictive list would state the epistemic settings relevant for the particular investigation. ‘Settings’ is thus an external notion; the world's contribution to a successful epistemic investigation, as it were. However, a cognitive agent who forms a belief based on a particular bit of evidence will, as a matter of course, believe the world to be co-operative in this sense, consciously or subconsciously.

continuously confronted with all sorts of raw uncensored evidence, ‘First-order’ evidence, which provides for an, at least initial, reason to hold a particular belief. But in addition we possess second-order beliefs in virtue of which we access and rank our first-order evidence. We take ourselves to have some correct opinions regarding the reliability of various pieces of first-order evidence. Indeed, the world would be an awful place to be if we didn’t. This is very clearly illustrated in Douglas Adams’ science fiction novel *Dirk Gently’s Holistic Detective Agency*. Adams sets his story in a future society in which robots have taken over all the dull and laborious functions in life. Robots do all the monotonous labour at the factories. They do the dishwashing, the car washing etc., etc. Robots have even taken over the television watching (the most advanced of the tele-robots being capable of watching as much as fourteen channels simultaneously). The most recent, state of the art, development is a brand new robot, which has been devised to care for religious duties. These ‘electronic monks’ believe for you! And they believe everything they see or hear without any sense of critical reflection. One day, however, things go terribly wrong: by a mistake one of the electronic monks is swapped with a tele-watcher robot. After just 5 minutes the poor electronic monk suffers a severe nervous breakdown: believing fourteen distinct television channels is more than even a 160 MB electronic monk can capacitate (the 160 MB being calculated as sufficient capacity to cope with five (inconsistent) major religions, sixteen complete party political broadcasts and Ronald Reagan’s autobiography). Adam’s scenario trades on the fact that evidence is despairingly rarely unequivocal. Nearly every time we possess a warrant for believing some belief P we also possess some contradictory warrant for believing \neg P. For example, while testimony from some informant warrants us in believing she is in a particular state of mind—say cheerful—close observation of her behaviour warrants us in believing she is not. One of life’s major challenges is to assess the reliability of such contradictory warrants and to form beliefs accordingly.

Electronic monks are incapable of doing this. We, on the other hand, although not infallible in that respect, have some degree of competence. We are capable of deciding between contradictory warrants and, generally, of accessing all sorts of first-order warrants and ranking them mutually with respect to credibility. That is, our cognitive functioning is not restricted to merely gathering all sorts of evidence and more or less blindly forming beliefs accordingly. We are, in addition, capable of critically assessing such justification. Some evidence is considered amenable to all sorts of distortion. For example, when a television advertisement e.g. tells us a particular brand of washing powder by far exceeds its rivals in efficiency and gentleness we automatically make certain reservations regarding this information. We envisage several possibilities for error: circumstances under which the (default) evidence is non-veridical. Other sorts of evidence, on the other hand, are considered highly plausible. For example, were we to discover, by inspection, that the car isn't parked where we left it 2 hours ago we would tend to assess this piece of observational evidence as highly reliable. In this case we would, on most occasions, not even bother to consider possibilities of error: scenarios in which the evidence for some reason might be misleading.

At my home university I always took it as evident that one of my fellow students had borrowed a car if I saw her driving one. At my present university, however, seeing a fellow student driving a car is pretty good evidence that she owns one. And there are, presumably, universities at which the same evidence can be safely taken to indicate that a fellow student seen driving a car recently has stolen it. Settings just vary. And to get them right, and hence adjust our first-order evidence appropriately, is crucial to competent cognitive functioning.

Second-order beliefs can perform the function of adjusting first order evidence appropriately. And second-order evidence secures that the second order beliefs do not do so blindly. But what is second-order evidence? It can be understood as a grasp of one's own settings, i.e. an ability to determine whether the present setting is one in which a given piece of first-order default evidence is to be considered

reliable. In archery an apparently favourable trimming of the steering device might send the arrow a few inches above the target. If it does so repeatedly the skilled archer will learn a lesson from this: some factors external to the steering device, a strong head wind say, influences the arrow's curve towards the target. Hence, strong head wind must be accounted for by aiming a bit lower than usual. Likewise some apparently favourable observational evidence might suggest some person has a particular character—is magnanimous and honest, say—but beliefs to just that effect might, disappointingly, prove fallacious. If it happens repeatedly that such first-order evidence turns out to be fallacious in a particular social environment the skilled cogniser will learn the lesson and aim a bit lower when forming beliefs regarding other people's characters based on observable evidence. The skilled cogniser will thus learn to read his own settings more carefully in the future: when situated in a particular social environment, possibilities of error will be considered likely in the light of, otherwise reliable, evidence. Ultimately second-order evidence—competence to adjust one's first-order evidence to the appropriate settings—is thus based on inductive inferences from past experiences.

How radical a modification must such second-order evidence be capable of accounting for? Well, in archery we are inclined to say that every practitioner, even a Zen master, has his limitations. In a situation with gale force winds and in which the target is very small and moving at very high speed it is highly probably that even a Zen master would miss his target. And if he is so exceptionally skilled that he would hit even under those circumstances we could readily add some further potential modification, such as momentarily suspension of the gravitational forces, the target turning invisible etc. under which he certainly no longer would hit the target. But that, in itself, does not challenge his being an archery master. To be an archery master his manipulation of the steering devices only has to be so constituted that he will hit the target within a range of *reasonable* hypothetically modified situations. We demand of him that his manipulation of the steering device is sensitive to a reasonable range of device-external factors such as for

example the general wind conditions. But we do not expect of the archer that he should be capable of taking into consideration such far-fetched hypothetical changes in device-external conditions as e.g. the target being removed at last moment by a guardian angel.

Similarly, in the case of belief formation, we must admit that potential modifications of circumstances are conceivable as a result of which even the most promising justification will be misleading—even for a skilled cogniser who has got things right regarding his own settings. The clock case illustrates this point. Foucault's justification for his belief is, *prima facie*, strong. He has just consulted his clock. In addition, Foucault has a qualified second-order belief regarding this sort of evidence. He is aware of his own settings. He knows his clock is reliable and hence that consulting it provides a strong warrant for believing what time of day it is. He is like the archer who has read the wind-conditions correctly, and manipulated the steering device accordingly, but whose arrow has been lead astray by a sudden change in the wind (*cf.*: *this* time the clock isn't working properly), but, miraculously, which has been put back on the right course by another, likewise unpredictable, windfall (*cf.*: Foucault consulted the clock at *that* particular time) and thus eventually hits its target. I said Foucault is lucky. And he is. But he is, at the same time, extraordinarily unlucky. Although he has a pretty good grasp of his own settings, has a suitable second-order belief regarding the reliability of consulting-the-clock evidence, he is wrong about his own settings in this particular case. He has the second-order belief that his clock is working properly. But it is not.

First-order evidence of the sort available to Foucault, namely looking at the clock, does not amount to knowledge under conditions in which the relevant second-order beliefs are mistaken, i.e. the conditions described in, or similar to those in, the clock case. This is rather trivial. However, the reciprocal claim, that whenever the relevant second-order beliefs are veridical, any (first-order) evidence counts as knowledge, might shed some light on our initial question: what it is for a

warrant to be constituted in the same, or very similar manner, under slightly modified circumstances. An initial, although by no means exhaustive approach to this question might be the suggestion that a warrant for P is identically, or very similarly constituted, if it is acquired (i) in a context where the relevant second-order beliefs P' are held constant; and (ii) in a context where the truth-values of those P' are held fixed. According to this suggestion, had Foucault consulted his clock a couple of days earlier while it was still functioning properly, his evidence would not have been constituted the same way as in the original case: although (i) above would be satisfied, (ii) would not. (i) would be satisfied since Foucault would still have believed that his clock was working properly if he had consulted it a couple of days earlier. But (ii) would not be satisfied since the relevant second-order belief—the belief that the clock is working properly—actually would have been veridical if Foucault had consulted the clock a couple of days earlier. Likewise, if Foucault is to look at a clock in a museum for old clocks, his evidence is again not constituted in the same way as in the original case: this time because (i) is not satisfied, although (ii) is. This time (i) is not satisfied since Foucault would not assume that an old clock in a clock museum is working properly. But (ii) would be satisfied since a second-order belief to just that effect, as in the original case, would have been fallacious.

I said common sense has it that we always have opinions regarding our own settings, that we possess second-order beliefs, and, furthermore, that it is close to platitudinous that these second-order beliefs, by and large, are veridical. That, of course, is not to claim that all our second-order beliefs are infallible. Some of them, such as opinions on whether we happen to be at home or in a museum for old clocks, are presumably infallible according to common sense. Others, such as, the belief for example that our clock is working properly, might be assigned merely a high degree of probability. Although it is a matter of course that we know whether a familiar clock is reliable, it is still in conformity with common sense that error occasionally sneaks in. However, in either case the relevant second-

order belief is detectable, as it were. Indeed it is essential to the idea of a second-order belief containing such modal notions as 'reliable' that they must occasionally be fallacious, if only very rarely. And it is natural to suppose that it is in virtue of the fallibility of our second-order beliefs, and the resultant rare pitfalls, that talk about being right and wrong makes sense in the first place. Far from being an obstacle, occasional error regarding one's setting is, in this sense, rather conducive for knowledge acquisition. Justification need in any case not take account of *this* kind of setting variation in order to count as knowledge. It suffices that the settings have been properly read with a certain degree of probability (whatever exactly that degree may be). It is, in other words, not necessary to rule out every possibility of an error scenario.

However, another sort of circumstances is also conceivable which renders first-order evidence misleading. To your big surprise you discover your car is not parked where you left it a few hours earlier. Your second-order belief here is that you are not suffering any sort of mental or perceptual disorder, that the viewing conditions are normal, that you are capable of recognising your car if confronted with it etc. etc. Each one of these second-order beliefs might be fallacious: Your 'friends' might have played a trick on you and, in your absence, painted your car a different colour, so that you now don't recognise it as yours, or they might have covered it behind a cleverly painted screen, or they might presently be hypnotising you into merely believing your car has been removed. Each of these three error-scenarios differs from the clock case in that you don't consider yourself to have to rule them out 'with a certain degree of probability'. You don't consider yourself to have to rule these far-fetched possibilities out for the simple reason that, under normal circumstances, they don't need to be taken into consideration at all. Under normal considerations our second-order beliefs urge us to take almost any kind of (first-order) perceptual evidence to be infallible. However, skilled cogniser as you are, if it turns out, repeatedly, that your belief regarding your missing car is fallacious, or more plausibly, that your belief regarding your missing car proves

fallacious in certain specified circumstances (e.g. when your friends are having their lunch-break), then you will take heed—you will take such error-scenarios into account (or at least take them into account under certain specifiable circumstances). In this case you will become more alert to clues about your settings also when forming perceptual based beliefs about your car's whereabouts. Now the problem is that one of the pitfalls suggested above, the hypnotising scenario, is such that no clues would be available to you in virtue of which you could determine what your settings are, i.e. determine whether your astonishment about the missing car is caused by genuine perception or caused by your friends getting a kick out of hypnotising you again. Your relevant second-order beliefs would, accordingly, be completely missing. Thus, here the relevant second-order belief is non-detectable.

Granted, it is possible *subsequently* to reveal some experience as an instance of misperceiving. You come out of hypnosis to find you are surrounded by your friends who are falling about with laughter at the success of their prank—and realise you haven't been perceiving at all, but rather have been victim to your friends' unkind hypnosis joke. Or, more realistically (hopefully), you wake up in the morning relieved that the missing-car nightmare you just suffered really was a nightmare and as such can be forgotten about immediately without involving police or an insurance company. Do such cases not count as instances of revelation of misperceiving; and ought they, accordingly not to be taken into consideration whenever we (more or less spontaneously) take ourselves to be in genuine perceptual settings? In some particular circumstances, for example when we are lying in our bed, perception-like experiences do actually appear to be unreliable. The great difficulty here is that the clues we ordinarily can take as indicators of such a setting are inaccessible once we start misperceiving. Like so many other things in life: when we need them the most they are least available.

The classical modal theories of knowledge and justified belief have various resources to meet such challenges and in the following four chapters I shall have

occasion to discuss a few of them in some detail. Chapter II is an exposition and criticism of the two most influential theories within modal epistemology: those due to Fred Dretske and Robert Nozick. I examine their theoretical strength in the light of a number of classical difficulties within epistemology. Likewise I will survey the points of criticisms launched against these theories. In particular I consider criticism of the specification of epistemological methods; the counterfactual semantics underlying the theories; and, lastly, the various sort of counterexamples that, at least *prima facie*, appear to undermine the plausibility of these theories. I conclude that these difficulties are not necessarily ascribable to the core ideas underlying Dretske and Nozick's theories but rather are the result of their specific formulations—in particular their exact formulations in counterfactual terms. My core thought is that the theoretical edifices in Nozick and Dretske's theories, if comprehended in terms of dispositionality rather than in terms of counterfactual reliance, retain their theoretical strength while avoiding various pitfalls. In particular, recent work on dispositionality indicates that the taxonomy of counterexamples to the tracking theory precisely mirrors the taxonomy of counterexamples to the counterfactual analysis of dispositional properties generally. This indicates that the various sorts of classical counterexample to the tracking theory are precisely those which are to be expected when one, misguidedly, attempts to subject a—in nature—dispositional concept to counterfactual analysis.

These findings in Chapter II are used for a more detailed discussion of dispositionality in Chapter III. Drawing from results in recent philosophical projects which also trade on dispositionality I am led to the idea that it is feasible to develop a significant amendment of modal theories of knowledge and warranted beliefs in dispositional terms. In Chapter IV I focus on the particular difficulties accruing from a conditional analysis of dispositional properties, the instances of 'masked', 'mimicked' and 'altered' dispositions. It will be argued that some of these difficulties can be remedied once it is recognised that alleged

counterexamples of this kind to the conditional analysis have to meet certain metaphysical and semantic constraints, notoriously the constraints we will define in the two principles of Variable Strictness and Compositionality Strictness. In Chapter V I will examine the possibility of a dispositional theory of knowledge in the light of these two principles. The driving thought here will be that knowledge is the result of exercising a dispositional skill to form veracious beliefs—just like competent archery displays as the result of exercising a dispositional skill to hit the target. This theory will lead us to a revised diagnosis of the problem cases for the classical modal theories and, eventually, to a more sophisticated version of a dispositional epistemology. In chapter V I will also survey some of the virtues inherent in such a dispositional theory: its theoretical strength in combating scepticism, its capability to clearly diagnose the classical problem cases for theories of knowledge and, not least, its upshot in an alternative way of viewing the structure underlying the wealth of beliefs whose aggregate constitutes our world-picture. It will give content to the idea that there really is a genuine distinction to be drawn between first-order and second-order beliefs as sketched above, a distinction germane to Wittgenstein's between 'empirical' and 'grammatical' beliefs, but it is argued that—in the light of a dispositional theory of knowledge and warranted beliefs—the second-order beliefs are aetiological in nature: they are concerned with how our first-order beliefs originate and, in particular, that they originate from the implementation of some genuine, epistemological method, perception, say, as opposed to an epistemologically flawed method such as brain-in-vat simulation. Second-order beliefs so understood as aetiological in nature are opposed to the Wittgensteinian 'grammatical' beliefs which are intrinsically metaphysical claiming, typically, that there *is* an external world, that there *are* other mind, that there *is* a past etc.

In this introductory chapter we have made some preliminary remarks about the relation between belief, warrant and truth. I have hinted at some of the central themes in modal epistemology such as non-accidental warrants, modal identity

criteria for warrants (i.e. what it is for a particular warrant to be identically constituted in slightly modified circumstances); internal vs. external warrants; sensitivity to epistemic settings and, finally, the question of to what extent a given warrant must be capable of accommodating radical alteration in a setting in order for the warrant to count as bestowing knowledge. It is now time to proceed to a more stringent examination of the arguments offered by some influential contributors in the field of modal epistemology.

Chapter II

Modal Theories of Knowledge and Warranted Belief

2.1 Penetrability of Epistemic Operators

2.2 Conclusive Reason

2.3 Four Misgivings

2.4 Tracking

2.5 Knowing More by Knowing Less

2.6 Summary

Synopsis

The present chapter is devoted to a thorough discussion of modal theories of knowledge and warranted beliefs. As mentioned, the development of modal epistemology has, more than anything else, been driven by the ambition to provide a satisfactory response to Edmund Gettier's challenge to the classical tripartite understanding of knowledge. There have (so far) been three stages in this development. When Gettier published his paper in 1963 it immediately triggered a wide variety of what we might phrase 'first-generation' responses. Among these were various versions of causal theories, no-false-assumption theories and indefeasibility theories. However, despite ingenious attempts to adjust and refine, none of these theories were sustainable. An important breakthrough came in 1981 with Robert Nozick's publication of his version of the so-called 'tracking theory'.⁴ It is not exaggerated to talk of a 'tracking turn' within epistemology analogous to

⁴ Although Fred Dretske, as it will emerge in 2.1 and 2.2, published what is essentially the same version of a tracking theory ten years earlier

the linguistic turn initiated by Frege and Wittgenstein's meaning-theoretic works. With Nozick the discussion got a new direction and a new focus. It founded, in other words, the second-generation of responses to Gettier. The transition from second- to third-generation has been less dramatic. It has not been initiated by any one particular ingenious piece of work. The second-generation discussions have just gradually passed out—not least due to frustration and dissatisfaction with some of the core ingredients in the tracking theory such as subjunctive semantics and method individuation. Alternatively epistemologists have gradually started to explore new modal fields such as formal information theory and various social-pragmatic approaches.

The present chapter is only concerned with second- generation approaches. The reason is that it contains philosophical insights which, when fully developed, are far more fertile than the various theories under the first- and third-generation paradigms. We believe, in other words, that it is worth to attempt a revival of the second- generation. That there is a hitherto ignored theoretical potential which can be extracted from these discussions and, indeed, ought to be fully developed before we get tired out and pursue new theoretical terrain.

The discussions proceeds by way of exposition (2.1 and 2.2), defence (2.3), comparison (2.4) and criticism (2.5) of tracking theories. 2.1 and 2.2 are expositions of two cardinal components in Fred Dretske's pioneering development of a tracking theory. In 2.3 we consider and dismiss an influential set of criticisms against this Dretskean theory put forth by Michael Williams. In 2.4 we present Robert Nozick's version of a tracking theory. There is a great deal of affinity between the two theories of Dretske and Nozick. However, the intent with the comparison is to highlight two respects in which they *diverge*. It will be argued that each of these divergences reveals disagreement on a deeper philosophical level: issues to do with subjunctive semantics and individuation of an epistemic method, respectively. Finally, in 2.5, we will subject tracking theories to a thoroughly critical analysis. We will focus on two major points: i) the various

categories of counterexamples to the theories; and ii) the highly counterintuitive consequences it has when one, with tracking theories, give up the Closure principle for knowledge.

2.1 Penetrability of Epistemic Operators

Fred Dretske, in his influential papers ‘Epistemic Operators’ (1970) and ‘Conclusive Reasons’ (1971), took the first steps in pioneering a modal theory of knowledge and justified beliefs; an approach which is now known as ‘tracking theory’. Dretske’s reasoning takes as its outset the apparently harmless observation that epistemic operators, or E-operators for short, such as ‘knows that’, ‘discovers that’, ‘has reason to believe that’ and ‘in the light of the given evidence it is probable that’ are not always closed under entailment⁵, or, to adapt Dretske’s terminology, do not always *penetrate* through entailments. That is, even if Q is a consequence of P and the epistemic operator E applies to P, this, in itself, is no guarantee that E also applies to Q. Or, more formally, the epistemic inference principle of *Closure*:

Closure

$$\frac{E(P), (P \rightarrow Q)}{E(Q)}$$

is not universally valid. An epistemic agent can thus E P, and know that P implies Q, and yet fail to E Q.

⁵ Dretske discusses whether E-ing is closed under *known* entailment. However, since any instance of entailment is knowable, any E-ing is closable under entailment if E-ing is closed under known entailment. We will therefore ignore the question whether the entailment in question really is known and instead constrain our discussion to the idealised instances where the knowable has become known.

Some sentential operators—such as ‘It is possible that’, ‘It is a fact that’ and ‘It is necessary that’—clearly penetrate through any instance of necessary consequence. They are *fully penetrating* operators. Other sentential operators such as ‘It is extraordinary that’ and ‘It is reproachable that’, on the other hand, fail to penetrate even through the most elementary logical consequences, as witness these examples:

(A) I bumped into my old school mate at a conference in New York;

entails

(B) I bumped into someone in New York

While (B) is a logical consequence of (A), and (A) is quite extraordinary, (B) isn’t remarkable at all. Likewise in the following, (A) entails (B), but while (A) is morally reproachable, the consequence (B) (arguably) is not:

(A) Jones left the party by car completely plastered,

entails

(B) Jones was drunk after the party

Sentential operators belonging to this category—‘it is extraordinary that’, ‘it is a mistake that’ etc.—are thus *non-penetrating* operators.

According to Dretske, Epistemic E-operators lie somewhere between fully penetrating and non-penetrating operators. Unlike non-penetrating operators, E-operators do penetrate the two entailments just mentioned and, generally, they penetrate through most elementary logical consequences. However, unlike fully

penetrating operators, E-operators fail to penetrate *all* logical entailments. E-operators are thus *semi-penetrating*. Generally they comply with Closure; but not always.

Dretske offers two distinct categories of cases in order to illustrate the circumstances under which the E-operators fail to penetrate through logical entailment. The first, and, as we shall see, least interesting category invariably involves some contextual shift of emphasis over the entailment. Here is an example:

(A) The roses are wilted;

entails

(B) It is *roses* that are wilted⁶

Assume someone during a walk in the local botanical garden, has come to E (A) that the roses are wilted. It is most likely that this happens as a result of scarce epistemic effort. In that case the person might E (A) and yet be in the dark about (A)'s consequent (B) that it really is roses (as opposed to for instance plastic counterfeits) that appear wilted. For: "there are certain presuppositions associated with a statement. These presuppositions, although their truth is entailed by the truth of the statement, are not part of what is *operated on* when we operated on the statements with one of our epistemic operators." (Dretske 1970, p. 1014; Italics in the original).

It is true, as Dretske points out, that there are certain presuppositions which are associated with any statement (what one might call 'tacit' information). If someone informs us that John murdered Bill, it will, in most cases, be received as

⁶ Dretske 1970, p. 1013.

a piece of information concerning the identity of the murderer. It was *John* who did it. However, the same sentence could be intended to inform us that the victim, Bill, not merely was severely wounded and became unconscious but actually was *murdered*; or, for completeness, that the victim now has been identified. It was *Bill* that John killed (Dretske 1970, p. 1023). The crucial point to notice is that in addition to whatever *is* being conveyed by a sentence such as ‘John murdered Bill’, something else is being tacitly assumed⁷. And although an informant may be in a suitable epistemic situation relative to the relevant piece of information, typically she will not be so suitably situated relative to the tray of tacit assumptions on which the relevant piece of information is served. An informant might thus be absolutely certain that *John* killed Bill, but, if put under pressure, less certain whether John *killed* Bill.

However, such cases of change of emphasis seem to indicate merely a far greater general contextual ambiguity than intended by Dretske. Dretske intended to disclose the semi-penetrability of E-operators in particular whereas the contextual penetrability failures under consideration apparently undermine the penetrability of *any* sentential operator—even one of the most penetrability stable operators such as ‘it is necessary that ...’ which Dretske, at the outset, considered to be fully penetrating and, indeed, introduced with the specific aim to contrast epistemic operators against . Consider:

- (A) Two apricots plus two mangoes give four pieces of fruit;
which entails
- (B) Two *apricots* plus two *mangoes* give four pieces of fruit.

⁷ Granted, someone may on a particular occasion attempt to communicate the full information:

John murdered Bill. But that is the exception.

(A), which states the mathematical fact that two and two equals four, is necessarily true. The necessary consequent (B), on the other hand, states the contingently true empirical fact that some particular fruit bowl has a certain composition: the bowl consists of two apricots and two mangoes—as opposed to, say, two bananas and two apples. Penetration failures trading on contextual factors therefore do not reveal anything about E-operators in particular.

In fact, even granted that a given sentence conveys different information depending on the contexts in which it is uttered, this fact in itself does not serve as a demonstration that E-operators fail to penetrate across logical *entailment*. For, if the propositional content of a given sentence really does swap midway through the ‘entailment’ due to change of emphasis, the logical status of that ‘entailment’ would, thereby, be rendered rather feeble. But if this is so, then there is no entailment which the E-operator could fail to penetrate through in the first place. Although ‘two apricots plus two mangoes give four pieces of fruit’ certainly entails ‘Two apricots plus two mangoes give four pieces of fruit’, it is not at all obvious that the contextualized:

(A) *Two* apricots plus *two* mangoes give *four* pieces of fruits;

entails the differently emphasised

(B) Two *apricots* plus two *mangoes* give four pieces of fruits.

If, on the other hand, one insists that (A) and (B) are equivalent then it demands the same epistemological effort to gain a warrant for the belief that (B) holds true as it does to gain a warrant that (A) holds true. But in that case one’s warrant for believing (A) must also penetrate through to (B). In either case we can conclude that there is no robust entailment from (A) to (B) which E-operators—or any other

sentential operator for that matter—fail to penetrate through. And, to be sure, the same considerations apply to Dretske's own example:

- (A) The roses are wilted
- (B) It is *roses* that are wilted

This leads us to Dretske's second, and far more interesting, category of cases illustrating the mere semi-penetrability of E-operators. This category encompasses cases of so-called 'contrast consequences': inferences of the following logical form:

$$P \rightarrow \neg(Q \ \& \ H)$$

where Q is incompatible with P and H is some hypothesis relative to which any evidence for P transforms into equally strong evidence for Q. Dretske's classical illustration of a contrast consequence is given by the Zebra case in which Stripe, questioned by his son, answers that:

- (P) The animals in the pen in front are zebras.

It follows that it is not the case that:

- (Q) The animals in the pen in front are mules.

In particular the animals in the pen are not mules that happen to resemble zebras because:

(H) The zoo authorities have cleverly disguised the zoo mules to look like zebras.⁸

Or, to mention another of Dretske's similarly structured examples: one comes to believe that:

(P) A nearby wall is red

after having paid careful attention to it. It follows that it is not the case that:

(Q) That nearby wall is white.

and, in particular it is not the case it is a white wall which happens to appear red because:

(H) The wall is being illuminated by trick light—special ultraviolet lighting—which makes all white things appear red.

However, the logical structure is more significant than the particular cases. What one should note is that for virtually any empirical P, a corresponding contrast consequence $\neg(Q \ \& \ H)$ can be constructed where Q is some proposition inconsistent with P and H is stipulation of some unusual epistemic conditions,

⁸ Dretske 1970, p. 1016. Compare Wittgenstein 1979, §588 "In a zoo there might be a notice 'this is a zebra'; but never 'I know that this is a zebra'"(!)

realisation of which would transform the relevant P-evidence into equally strong evidence for Q.

To return to the Zebra case, almost all evidence the conscientious Stripe can adduce when challenged by his son as to whether he really E's P will, merely in virtue of the logical structure of the case, be perfectly consistent with (Q & H). The fact that Stripe E's P does therefore not, in itself, secure that he E's $\neg(Q \& H)$. In order to E $\neg(Q \& H)$ he must in addition be in possession of *independent* evidence in favour of $\neg(Q \& H)$ which, in turn, only seems practical if he is in possession of evidence against H. If, as Dretske assumes, the father does not possess any independent epistemic warrant against H, then we must conclude that his E-ing of P fails to penetrate to the contrast consequence $\neg(Q \& H)$.

We said that evidence for some proposition P is gathered on an 'invisible tray' of collateral information. What this second category of examples brings out is that this tray of tacit assumptions is not only made up of implicit information which, strictly speaking, is extractable from P. This tray can, in addition, include parcels of information which, even strictly speaking, are entirely independent of any informational content extractable from the sentence conveyed. It is this characteristic that demarcates these cases in this second category from the ones trading on contextually generated ambiguities. If a botanist claims that the roses are wilted, the tacit assumption that it really is roses which are wilted can, strictly speaking, be deduced from the explicit claim that the roses are wilted. If, on the other hand, Stripe claims that there is a zebra in the pen, the tacit assumption that the zoo authorities have not cleverly disguised the zoo mules cannot, even strictly speaking, be deduced from what has been explicitly stated. We can say that the implicit assumptions featuring in our contrast-consequences are *content-neutral* whereas the implicit assumptions featuring in the cases of contextual penetrability failure are content-bound.

At this point it might well be objected that Stripe *does* E $\neg H$. A lot can be said in favour of that - in particular if E is interpreted as 'in the light of S's evidence, it is

probable that'.⁹ Indeed, Dretske's viewpoint in subsequent discussions of these matters seems to favour probability considerations as being the ultimate arbiter of whether an agent E's subject matters such as $\neg H$. (Dretske 1981, pp. 130-1). The thought being, roughly, that an epistemic agent E's a contrast consequence such as $\neg(Q \ \& \ H)$ iff the objective probability for H is sufficiently small. (We shall return to this suggestion and discuss it in some detail in Chapter IV and Chapter V). Furthermore, in 'Epistemic Operators' Dretske argues against the idea that one is, as a matter of fact, warranted in believing the relevant $\neg H$. He writes:

The fact is that we habitually take such matters for granted, and although we normally have *good* reason for making such assumptions, I do not think these reasons are sufficiently good, not without special precautionary checks in the particular case, to say of the particular situation we are in that we *know* conditions are normal. (Dretske 1970, p. 1015; italics in the original)

It is no easy matter to interpret what exactly 'good reason' is supposed to mean here and, in particular, what standing 'good reasons' are supposed to have in the light of the other E-operators that Dretske discusses. For now it will suffice to note that even if, as against Dretske's claim, one is warranted in believing the relevant $\neg H$, the warrant in question will be of a different nature from the warrant one has for believing the corresponding P-proposition. If we, for example, grant that Stripe in the Zebra case holds *some* warrant for his belief that $\neg H$, that the zoo animals are not clever fakes, then his warrant for this belief must be distinct from the warrant he has for holding that P, the creatures in the pen in front of him are zebras. His warrant for the latter is based on direct perceptual observation, whereas his 'good reasons' for the latter, whatever exactly they may be, must be based on something distinct from direct perceptual evidence. So, even granted E-penetrability, the warrant W in virtue of which Stripe E's the antecedent P is a different one from the warrant W' in virtue of which he E's the contrast consequent $\neg(Q \ \& \ H)$. Accordingly the warrant for P does not *transmit* to $\neg(Q \ \& \ H)$.

⁹ And, indeed, a lot has been said to just that effect. See Wright 1997 and 1998b.

H). That is, even if the E-operator penetrates through the entailment, Stripe does not come to E the consequent *in virtue of* the same warrant by which he comes to E the antecedent. Recently, Wright has elaborated on this distinction between penetrability and transmission of E-operators and, in particular, which consequences it has, or ought to have, for a proper treatment of scepticism.¹⁰ We shall return to the issue in Chapter V.

2.2 Conclusive Reason

We will proceed under the assumption that Dretske is right in his claim that epistemic operators are merely semi-penetrating. This claim is, on the face of it, rather harmless; however, an array of significant epistemological insights can be extracted from it.¹¹ Firstly, it reveals that a Dretskean epistemologist—to the extent that she believes that the concept of knowledge involves one of the E-operators discussed above—is committed to an externalist view of knowledge where we by ‘externalism’ will understand the following view:

Epistemological Externalism

Some of the factors needed for a belief to be warranted for some agent are not cognitively accessible to that agent.

Consider the paradigmatic illustration of penetration-failure stated above. Situated before the zebra’s pen in a zoo one knows P, that one is confronted with a zebra, but one fails to know the contrast consequence $\neg(Q \ \& \ H)$ that one is not

¹⁰ Wright 1997 and 1998b. In Wright 1998b, considerations particularly to do with transmissibility of E-operators are also applied to shed light on certain paradoxes which arise once one endorses semantic externalism.

¹¹ And have indeed been so extracted by Dretske himself in his significant, but widely ignored, paper, ‘Conclusive Reasons’ (1971).

confronted with a cleverly disguised mule. The reason why one fails to know the latter, the thought is, is that one fails to E, and thus know, that the hypothesis, H, of this particular contrast consequence is false—i.e., the creature in question is a clever fake. So knowledge that $\neg(Q \ \& \ H)$ is not a necessary condition for knowledge of P. However, the *truth* of $\neg(Q \ \& \ H)$ is still a necessary condition for knowledge of P, since it is a necessary condition for the *truth* of P. (And the truth of P is, in turn, a necessary condition for knowledge of P). Hence, at least one essential aspect of the epistemic situation characteristic for knowledge is thought by Dretske to be purely external, *viz.*, the fact that certain preconditions as a matter of fact prevail. As long as these preconditions as a matter of fact are satisfied, one can, according to Dretske, know anything that presupposes them. It is the fact that these presuppositions hold, not our E-ing them, which is significant when evaluating a given knowledge claim. Dretske is, on numerous occasions, explicit about this: knowing that P does not entail that one also knows Q, some necessary conditions for coming to know that P; nor does knowing that P entail that one knows that one knows P.¹² We will follow prevailing terminology and call this latter principle *Iterativity*:

Iterativity

$$K(P) \rightarrow K(K(P))$$

Iterativity has often been associated with epistemic internalism—the view that all relevant factors involved in an agent’s knowing that P must be ‘open to view’ to that agent. It is easy to see why one would associate internalism with Iterativity: if an agent knows that she knows P whenever she knows P she must also know that all relevant factors involved in knowing that P obtain whenever they do. However, unless it is already settled that knowledge implies transparency of the knowledge-conferring state of affairs, i.e. unless internalism already is presupposed, there is

¹² See in particular, Dretske 1971, pp 17-19.

no reason to suppose that knowing that all the knowledge-conferring states of affairs obtain also lays these knowledge-conferring states of affairs open to view. Generally, there is no reason to think the outermost knowledge operator in Iterativity lays anything open to view that is not already laid open to view by the innermost knowledge operator. Hence, Iterativity is perfectly consistent with epistemic externalism. It may so happen that the non-transparent states of affairs that the externalist takes to constitute knowledge is always accompanied by other non-transparent state of affairs which, according to the same externalist standards, constitute knowledge that one knows.¹³

This much, however, is true: *negation* of Iterativity is sufficient for externalism. Iterativity is not demarcating for internalism:

Epistemological Internalism

All of the factors needed for a belief to be warranted for some agent must be cognitively accessible to that agent.¹⁴

since Iterativity also holds for certain, although not for all, versions of externalism. Iterativity does however hold for *all* plausible versions of internalism¹⁵ and an explicit *rejection* of Iterativity, such as that by Dretske, is thus a demarcating characteristic for externalism. Internalism thus implies Iterativity, but not *vice versa*.

As Dretske's considerations over semi-penetrability bring out, one can conceive of contrast consequences for virtually all empirical propositions. Furthermore, contrast consequences whose realisation implies knowledge failure even in the

¹³ As pointed out in Wright 1993c.

¹⁴ For a helpful discussion of other versions of internalism and their mutual relations, see Dancy (1985), pp. 131-5.

¹⁵ There are (plausible) epistemological positions advertised under the heading 'internalism' which denies Iterativity. However, these are only internalistic in the very minimal sense that they require *some* internal component to be incorporated in an otherwise externalistic theory.

light of the most favourable evidence. Put differently, even the most favourable evidence depends for its power to transform a belief into knowledge on certain collateral states of affairs, states of affairs to which usually one is completely blind, but whose significance suddenly is made vivid by the fabrication of Dretskean contrast consequences. Dretske's considerations thus illustrate what has been phrased 'holism of justification' or, in Wright's terminology, 'information-dependence' of evidence: the view that any piece of evidence maintains its supportive status only relative to a network of background beliefs.¹⁶ It is precisely this tacit background information that is challenged by the various contrast consequences and their constituent H-hypotheses. The H-hypothesis can either call into question some local background information, as in the Zebra case, or it can challenge a background information of a more global character, as is the case in the familiar Cartesian-inspired sort of contrast consequences: you believe you are sitting performing one of your numerous tedious duties by reading this text, but does your evidence supporting this belief rule out that you are merely suffering one of your usual nightmares and shall soon wake up to a blessed world void of such duties? The point mentioned above, that the existence of non-penetrable contrast consequences implies epistemic externalism, in this way provides a potential knock-down argument across the board against epistemic internalism regarding knowledge: for virtually any knowable P we can point to one relevant aspect of the epistemic situation, namely the absence of contrast consequence, which never will be accessible to the epistemic agent. Hence internalism does not hold for virtually all P's.

¹⁶ Information-dependence has been a popular topic among philosophers, not least since the publication of Wittgenstein's notes 'On Certainty' (Wittgenstein 1979). For a (limited) selection of illuminating discussions, see Marie McGinn 1989, Williams 1996, Ch. 2, Davies 1999; and Wright 1986, 1991b, 1997, 1998b.

Note, though, that this prospect, a knock-down argument against epistemic internalism, will be conditional on two crucial factors:

- (i) that one is, as a matter of fact, debarred from E-ing when it comes to tacit, but essential background beliefs; and
- (ii) that the consequential lack of warrant for the background beliefs does not, in turn, undermine the warrants for more familiar empirical beliefs

It might be that Dretske is wrong in his assumption that it never is transparent to the agent whether or not some contrast consequence has materialised (i) and, even if that assumption is correct, it still remains to be settled whether this fact undermines knowledge claims for more mundane beliefs (ii).

Dretske's discussions in 'Epistemic Operators' reveal a significant epistemological insight, *viz.*, that warrants for standard empirical claims are useless if not collected on invisible trays of tacit background information. Furthermore, it appears that some of this background information may itself transcend any means of justification. But this finding is, in itself, as much (if not more so) an argument for global scepticism as it is an argument for semi-penetrability: granted (i), one could as well proceed, due to a strong conviction in penetrability, to deny (ii). In any case this appears, *prima facie*, to be as natural a line to choose as Dretske's, *viz.*, guided by a strong conviction in (ii), to proceed by denying penetrability. It is, more than anything else, support for (ii) that is Dretske's main objective in his follow-up paper 'Conclusive Reasons'

His starting point here is the classical epistemological question: which relation has to hold between a belief and the believed state of affairs in order for the belief to qualify as knowledge? His suggestion is that one must possess some evidence (R) for one's belief that serves as a 'conclusive reason' where conclusive reasons are characterised as follows:

Conclusive Reasons

The evidence R constitutes conclusive reasons for some epistemic agent S for believing P iff S would not have had the evidence R for P unless P had been the case.¹⁷ ($\neg P \Box \rightarrow \neg(S \text{ possesses } R)$).¹⁸

With the further assumption that:

(Dii) S believes that P (solely) on the basis of her evidence R,¹⁹

it follows that:

(Diii) $\neg P \Box \rightarrow \neg(S \text{ believes } P)$.

The conjunction of (Dii) and (Diii) constitutes in this way what Dretske takes to be necessary, and sufficient for S to know P²⁰.

It is worth noticing that Dretske uses R, 'conclusive reasons', in a rather technical sense. For a start there need not be any reasoning at all involved in having a 'conclusive reasons'. Dretske speaks of R as the 'epistemic credential' for holding P (1971, p. 1) or simply as the 'state of affairs' to which the epistemic

¹⁷ Dretske 1971, p. 12. The same counterfactual relation is featured in Dretske 1970.

¹⁸ We use ' $P \Box \rightarrow Q$ ' as a formal representation of the counterfactual relation: were P to be the case, then Q would be the case.

¹⁹ Formulated in Dretske 1971, p. 1.

²⁰ Dretske 1971, pp. 12-3. Dretske formulates the further requirement that R either be an experiential state for which it does not make sense for S to doubt that R or, if it does make sense to doubt R, S must know R to be the case. This further requirement is not significant for our present purposes and, for the sake of simplicity, we shall ignore it in what follows.

agent's conviction can be traced (*Ibid.* p. 13) where he takes 'states of affairs' to embrace both evidence, grounds and reasons (*Ibid.* p. 1). As Dretske writes:

I do not wish to suggest by my use of the word 'reason' that when S has conclusive reasons for believing P, S has *reasoned* his way to the conclusion that P is the case from premises involving R or that S has consciously used R as a reason in arriving at the belief that P. I am inclined to think (but shall not now argue it) that when one knows that P, on whatever basis this might be, little or no reasoning is involved. (*Ibid.*, p. 15; original italics)

These rather loose characterisations indicate that Dretske would have appreciated the availability of some theory-neutral and all-embracing epistemic notion. We shall henceforth refer to R as an agent's *warrant* for holding her belief.

Why should (Dii) and (Diii) constitute an exhaustive, and correct analysis of knowledge? Dretske approaches this question first with an array of cases where (Dii) fails and points out, in each particular case, that whatever factor it is that undermines (Dii), precisely that factor is, intuitively, the reason we would like to give when explaining why the epistemic agent in question lacks knowledge. Some agent S knows, for example:

(P) that the child's temperature is normal (37.3°C), on the basis of,

(R) her reading of the thermometer which she has just removed from the child's mouth

(Diii) applied to this case yields:

(Diii) $\neg P \Box \rightarrow \neg(S \text{ believes } P \text{ on the basis of reading the thermometer})$ ²¹

²¹(Diii) is ambiguous between (a) $R \Box \rightarrow (\neg P \Box \rightarrow \neg(S \text{ believes } P))$, (b) $R \Box \rightarrow (\neg P \Box \rightarrow \neg(S \text{ believes } P))$ and (c) $(R \ \& \ \neg P) \Box \rightarrow \neg(S \text{ believes } P)$. Here and throughout we read it as (c).

What could falsify (Diii)? All the candidates that spring to mind—that S would believe P despite $\neg P$ because the thermometer would have stuck at 37.3°C anyhow, or S would believe P despite $\neg P$ because the child is a notorious cheater who would have manipulated the thermometer to show 37.3°C (e.g., by secreting an ice-cube in her mouth)—all these reasons are precisely the reasons we put forward when giving expression to our intuition that the agent, in these cases, fails to know—although, luckily, may have formed a true belief on this particular occasion.

Dretske then proceeds to some more systematic arguments in favour of his account. He points out that it possesses the theoretical resources to deal satisfactorily with the Gettier cases.²² Take, for instance, the Gettier case mentioned in Chapter I, that is, the Clock case. Foucault looks at his stopped watch and concludes, truly, that it is now 6 o'clock. Foucault thus forms a true warranted belief which under the conditions described, nevertheless does not qualify as knowledge. The reason, Dretske tells us, is that (Diii) is not satisfied in this case: had it not been 6 o'clock, but five past, say, Foucault would still have believed it was 6 o'clock on the basis of consulting his (stopped) watch.²³

Furthermore, Dretske points out, his theory steers clear of the drawbacks inherent in such modal theories of knowledge as reliability theories and causal theories—the major modal theories developed in response to the Gettier challenge.

For example, a reliability theory falls prey to the so-called 'Lottery Puzzle':

²² Dretske 1971, p.13 In Hetherington 1998 it is argued that no remedy is called for, i.e., that, on reflection, Gettier cases are not genuine counterexamples to the classical tripartite definition of knowledge. This is, to put it mildly, a controversial claim. And, furthermore, it is a claim in desperate need of (better) argumentative support.

²³ Although Dretske tells his story on the background of a slightly different case. Dretske 1971, p. 13, n.13.

Lottery Puzzle

Gambler buys a ticket in a lottery where the chances of winning are 1 in a 1,000,000. When put under pressure, Gambler admits that he honestly believes (P), that his ticket will lose. Gambler has the best possible evidence for believing P (in terms of probability). Now assume that P actually is the case, that Gambler does have a losing ticket. Why are we still hesitant to ascribe knowledge of P to Gambler? Indeed, why are we more likely to ascribe knowledge of P to Gambler *after* he has come to learn P by reading a newspaper? After all, the chances are far more than 1 in a 1,000,000 that the newspaper, due to for instance a misprint, might be inaccurate regarding P. Reading the newspaper appears thus not to provide Gambler any better evidence for P than the (probabilistic evidence he already possessed).²⁴

Obviously a reliabilist theory for knowledge cannot account for this puzzle. Dretske, however, points out that it is the (Diii) clause that creates the crucial asymmetry between plain probabilistic evidence and testimonial evidence issuing from consulting the newspaper (Dretske 1971, pp. 3-4): Were $\neg P$ to be the case, were Gambler as a matter of fact to possess a winning ticket, he would still believe he did not if his warrant consisted entirely in general probabilistic considerations. These general considerations are completely insensitive to the truth of P—they would issue in exactly the same warrant for P in a P-scenario as they would in a $\neg P$ -scenario. On the other hand, the testimonial warrant gained by reading the newspaper is, if not infallible, then at least sensitive to the truth of P in the manner (Diii) requires: were Gambler to possess a winning ticket, he would come to believe just that by consulting the newspaper. The warrant that one gains from reading a newspaper is sensitive to the truth of P in the sense that the information it offers, although frequently distorted and occasionally even just plain false, is at least to some degree hooked up with the facts: newspaper information is the result of an outgoing investigation specifically aimed at truth. In

²⁴ Paraphrased from Harman 1968. In Harman's story the lottery evidence is contrasted to testimonial evidence generally. The twist, that the testimonial evidence in question is the lottery section in the newspaper, is added in DeRose 1995. The Lottery puzzle was first introduced in Kyburg 1961.

that sense it makes a difference whether P is true or false. Were P to be false, the newspaper would report its falsity; not with 100 (or 99.9999) per cent certainty, but at least with some acceptable degree of probability. This goes against the general probability considerations which would get things right with an unacceptable 0.0001 per cent probability, were P to be false.

Dretske also shows his own theory to be superior to a causal theory of knowledge. He illustrates this by the Lava case:

Lava Case

Ben has come to believe (P) that the mountain M erupted many years ago. The basis for his belief is the presence of solidified lava throughout the countryside surrounding M, solidified lava, furthermore, which, correctly, can be causally traced back to P. However, not far from M is another mountain N. The geology of the area is such that, at the point in time at which M erupted, the probability that N should have erupted was much higher than the chance that M should have erupted. Furthermore, had N erupted it would have prevented M from doing so. Finally, had N erupted, the distribution of lava would have been very similar to the actual formation in the countryside—and sufficiently similar for Ben to fail to notice any relevant differences.²⁵

Intuitively Ben does not know P in this case since it is luck, far more than epistemic skill, that ensures that he holds a true belief. However, a causal theorist would be obliged to ascribe knowledge to him. Dretske, on the other hand, can point out that his (Diii) is not satisfied: Had M never erupted, Ben would, as against (Diii), still have believed it did, since N, in that case, would have erupted and thus provided evidence which Ben still would have interpreted as evidence supporting P

Having shown the intuitive appeal of his theory and having argued for its superiority over competing theories, Dretske goes on to give a little more substance to the content of (Diii). To recapture: in 'Epistemic Operators' Dretske established

²⁵ Reconstruction from Dretske, *op. cit.*, pp. 4-5.

- (i) that one is, as a matter of fact, debarred from E-ing when it comes to tacit, but essential background beliefs; and

However, in order in addition to render plausible

- (ii) that the consequential lack of warrant for the background beliefs does not, in turn, undermine our E-ing of more familiar empirical beliefs

Dretske owes us an *independent* account of what it as a matter of fact takes to E.

With a contentious account of what it takes to satisfy (Diii), and thus an independent account of what it as a matter of fact takes to know, Dretske provides precisely that.²⁶ So, given that his attempt will be successful, abandoning (global) Closure is the only reasonable upshot. And whether it will be successful depends to a large extent on whether Dretske can provide a plausible semantics for the subjunctive claim featuring in (Diii)—in particular an account of which nearby $\neg P$ scenarios one ought to take into consideration under this evaluation. Obviously it will not do to identify relevant nearby scenarios with those scenarios where the background beliefs in question as a matter of fact hold true. In that case he would have offered a mere *ad hoc* justification for (ii) above.

So, which kind of hypothetical situations ought one, according to Dretske, to take into account when evaluating the subjunctive conditional featuring in (Diii)?

Dretske writes:

Statements such as [(Diii)], then, even when R and P are expressions for particular states of affairs, express a general uniformity, but this general uniformity is not that whenever a state similar to R is the case, then a state

²⁶ It is interesting to notice that Dretske, in this project, has constrained himself to considerations to do with *knowledge* rather than E-operators generally. But, naturally, the prospect of the announced result certainly makes the project well worth to pursue even if, at the end of the day, the result is applicable only for the concept of knowledge.

similar to P will also be (or have been) the case. The uniformity in question concerns the relationship between states similar to R and P *under a fixed set of circumstances*. Whenever (a state such as) R *in circumstances C* then (a state such as) P where the circumstances C are defined in terms of those circumstances which actually prevail on the occasion of R and P. (*Ibid.*, p. 7; italics in the original)

Dretske further characterises C as those circumstances “prevailing on the occasion which are logically and causally independent of the state of affairs expressed by the *antecedent* of the conditional.” (*Ibid.*, p. 8; italics in the original) To illustrate, Dretske offers the example of Fate who midway through a poker hand remarks that if his neighbour had not folded he (Fate) would have been dealt a royal flush. There, the counterfactual claim is meant to characterise the situation as it would have been had the neighbour not folded, while everything else remained fixed, that is, a situation with *this* particular distribution of cards in the remainder deck, *this* particular seating arrangement and *these* other players remaining in the game. It does not count to challenge Fate’s counterfactual claim by pointing out that had the neighbour not folded *but* some of the other players had, Fate would still not have been dealt a royal flush. Thus to introduce new counter parameters²⁷, such as the folding of a *de facto* non-folding player, is only allowed if this additional counter parameter is *relevant*, i.e. intimately connected, logically, causally or otherwise, with the original counterfactual antecedent. And whether an additional counter parameter is so connected, Dretske tells us, is determined entirely by features prevailing in the actual situation. There is a fact of the matter as to whether the collateral gambler is so wired psychologically that he would have folded, had Fate’s neighbour played on.

One important consequence of this view is that it is not contextually determined how strict are the standards a given piece of evidence has to comply with in order to qualify as knowledge. (Diii) is evaluated relative to actual features of the

²⁷ Introducing a counter parameter, in the present context, simply means strengthening of the counterfactual antecedent.

present situation—not relative to which context a knowledge-claim is uttered in.

Dretske writes:

It simply will not do to insist that in concerning himself with the possibility of mistake a sceptic is setting artificially high standards for knowledge and, therefore, may be ignored when considering ordinary knowledge claims. ‘You know that you are not dreaming’ does imply that [Diii] is satisfied, and it seems to me quite a legitimate line of argument for the sceptic to insist that if [Diii] is not satisfied, if you [despite your warrant to the contrary] might be dreaming or whatever, then ‘you know you are not dreaming’ is false. (*Ibid.*, p. 19)

The standards for epistemic warrant are thus determined *purely* by the range of fixed C-conditions. And these, in turn are determined by the counterfactual antecedent together with actual states of affairs. The more radical changes a counterfactual antecedent brings about, the narrower is the scope of fixed C-conditions. In a rather harmless antecedent such as “were Fate’s neighbour not to fold” virtually everything else, save the distribution of the rest of the deck, ought to be held fixed. In more radical antecedents such as “were Fate to be a brain in a vat” on the other hand, hardly anything can be held fixed: disturbingly many factors in Fate’s present setting are logically, causally, or otherwise hooked up with the fact that Fate is not a brain in a vat (BIV). Obviously, it becomes increasingly difficult to rule out error as more and more hypothetical situations are taken into consideration. Hence, Fate’s warrant, whatever exactly it consists in, might well endow Fate with the knowledge that his neighbour folded and yet be too weak to provide Fate with knowledge that he is not a BIV. Even though the latter is, arguably, a logical consequence of the former.

It appears thus Dretske has succeeded in developing a concept of knowledge according to which one can know a wide variety of mundane facts and yet fail to know about such fundamental matters as one’s own non-BIV-hood. In particular, one can know a variety of empirical facts without knowing the corresponding contrast consequences of these empirical facts. This is so because the H-

hypothesis featured in the contrast consequences has precisely the nature of a typical sceptical claim, the realisation of which is perfectly consistent with all available evidence. Hence, unless one can come up with good objections to this analysis, one is compelled to accept its two consequences: i) denial of the universal validity of Closure and ii) a degree of externalism regarding E-operators.

2.3 *Four Misgivings*

Michael Williams has launched a severe and influential criticism against this Dretskean theory (Williams 1996, pp. 330-6). In particular, against its inherent denial of Closure. Williams summarises his criticism as follows:

Dretske's case against Closure rests on four confusions: not distinguishing what I need to know to come to know a given proposition from what I know when I do know it (equating a rejection of epistemological priority with a denial of Closure); conflating the presuppositions of P with those of my knowing that P; using examples that work against Closure only if my knowing that P entails my knowing that I know that P, which no externalist is in a position to assert; and introducing, but then ignoring, scope distinctions in connection with epistemic operations [...] when these confusions are cleared away, nothing remains of Dretske's case for non Closure. (*Ibid.* pp. 335-6)

Regarding the first point of criticism—that Dretske equates what I need to know in order to come to know a given proposition with what I know when I do know it—Williams acknowledges the externalist aspect of Dretske's theory, that one can come to know an empirical proposition P even though one does not (first) know certain necessary conditions for P, such as the contrast consequence $\neg(Q \ \& \ H)$. What is essential is the *truth* of these presuppositions—not an agent's epistemic relation to them. But, Williams maintains, once P is known, the agent has gained a potential warrant in virtue of which she also can come to know the contrast consequence: (subsequently) the agent simply infers the contrast consequence $\neg(Q$

& H) from P. Inferring a necessary consequence from a known premise P is, Williams claims, consistent with Dretske's externalism:

If I do come to know that P, do I *thereby* also come to know that Q [the contrast consequence of P] (supposing that I know that P entails Q)? It is far from obvious that Dretske's, or any other, externalist account of knowledge requires answering "no" to this question. (*Ibid.* p. 332; my square-brackets, original *italics*)

There are, however, a number of good reasons why an epistemic agent cannot subsequently gain a warrant for the contrast consequence simply by inference from a known premise P. One is the reason already discussed, that independently of the status of Closure, Dretske's considerations certainly demonstrate that *Transmission* fails through inference to contrast consequences.²⁸ Another (closely related) good reason why one cannot subsequently gain knowledge of the contrast consequence by inference from P is that such an inference does not provide one with 'conclusive reasons' for believing the contrast-consequence: were the contrast consequence to be false, one would still believe it to be true on the basis of inferring it from P.

However, the best reason why one cannot subsequently come to know about the contrast consequent by inference from P is that such an inference, in the very least, presupposes Closure²⁹. But, given Closure, one never gets as far as knowing P. For the first step in Dretske's investigation was to demonstrate that one, as a

²⁸ The point is that $\neg H$ sometimes is part of the collateral information necessary for gathering evidence for P. When this is so one cannot, without circularity, transmit the evidence one has for P to the contrast consequence $\neg(Q \& H)$ and thus gain evidence for $\neg H$. For details, see Wright 1997. The difficulties connected with transmission of evidence to collateral beliefs is also discussed in Chapter V.

²⁹ Here I assume, for simplicity, that *Transmission* always implies *Closure*. For a more detailed discussion of the relation between the two principles, see Chapter V.

matter of fact, fails to establish any epistemological link, least of all knowledge, to the ‘invisible’ background beliefs essential to the acquisition of knowledge of more mundane beliefs. A natural reaction at that stage was to insist on Closure and thus give up the mundane knowledge claims in tandem. In fact, this move could only be prevented by Dretske’s supplying an independent account of knowledge according to which one does know the mundane empirical facts. Only in virtue of this independent account is Dretske allowed, in a non *ad hoc* manner, to opt for an alternative strategy, *viz.*, to give up Closure and thus preserve knowledge of ordinary empirical claims. Given this dialectical background, one cannot suddenly come back and claim full validity of Closure. A motivated insistence on Closure must involve one of the following two tasks: (i) demonstrating that Dretske’s finding from ‘Epistemic Operators’—that we are debarred from knowing contrast-consequences—is ill-founded, or (ii) demonstrating that Dretske’s theory of knowledge developed in ‘Conclusive Reasons’—a theory according to which we *do* know mundane, empirical facts—is fallacious. Simply insisting on Closure in a context in which Dretske has argued carefully against it is merely question-begging.

Indeed, the insight highlighted by Dretske that background beliefs, and, *a fortiori*, contrast consequence, often transcend warrantability renders Closure a highly unstable contention. With Closure there are two options: Either, as appears to be Williams’ suggestion, one can opt for Moorean *Modus Ponens* from knowledge of P to knowledge of $\neg(Q \ \& \ H)$; or one can, equally justifiably, make the sceptical *Modus Tollens* from the denial of knowledge that $\neg(Q \ \& \ H)$ to the denial of knowledge that P. Grant Moore and the sceptic the principle of Closure, and the two will enter a hollow but never-ending debate. It is naive, with Williams, to assume the Moorean *Ponens* will be the last word in such a dispute.

In any case, according to our reading, Transmission of knowledge of some P-belief to its contrast-consequence is an opportunity which is explicitly rejected by

Dretske. According to Dretske, knowledge of P cannot in this manner be transmitted through to the contrast consequence of P.

Williams' second point of criticism is that Dretske conflates the presuppositions of P with those of an agent's knowing that P. Williams focuses attention on Dretske's Wall case. In this example, an agent sees, and comes to know that:

(P) The wall is red.

The presupposition involved here, Williams rightly states, is the assumption that:

(¬H) The lighting is not a 'trick' lighting.

It is in virtue of ¬H that counter-possibilities such as:

(Q) The wall is (really) white

are precluded. Williams furthermore grants Dretske's point that the agent does not know ¬H. *But*, Williams continues, neither is ¬H a necessary consequence of P. ¬H is only a necessary consequence of the agents *knowing* that P. This is where Dretske makes his mistake, according to Williams: he confuses the entailment $P \rightarrow \neg H$ with the relevant entailment to be considered, which is:

(1) $K(P) \rightarrow (\neg H)$.

A counterexample to Closure of knowledge over *this* entailment would be a case where:

(2) $K(K(P)) \ \& \ \neg K(\neg H)$.

But, Williams points out, Dretske's example only establishes that:

(3) $K(P) \ \& \ \neg K(\neg H)$.

Hence, Williams argues, once this 'confusion' has been clarified, it can be seen that Dretske has not succeeded in producing a genuine counterexample to Closure. Williams concludes therefore that the Wall case example only has any weight against Closure by the further assumption of Iterativity which would render (2) and (3) equivalent—but, he points out, since Iterativity is inconsistent with externalism, Dretske cannot help himself to that principle. (Williams 1997, pp. 332-4).

Everything Williams says here is correct³⁰—but completely irrelevant to Dretske's case. Dretske only discusses failure of Closure over entailments from P to the *contrast consequences* of P, he never as much as mentions entailments from P to P's *presuppositions*³¹. What Williams describes as 'the legacy of Dretske's having confused the presuppositions of a given proposition with the presuppositions of my knowing that proposition' (Williams 1996, p. 334) is thus indeed the legacy of a confusion, but not a confusion on the part of Dretske. *Dretske's* counterexample to Closure accrues by considering the entailment:

(4) $P \rightarrow \neg(Q \ \& \ H)$,

where the agent in question knows P but lacks knowledge of the contrast consequence $\neg(Q \ \& \ H)$. A clear-cut counterexample to Closure and, furthermore,

³⁰ Save the last remark that Iterativity is demarcating for internalism.

³¹ Cf. e.g. Dretske 1970, p. 1019: "I now wish to argue that these operators do not penetrate to a certain set of contrast consequences" or, p. 1016, "But it does not follow from the fact that I know that the wall is red that I know that it is not white cleverly illuminated to look red."

a counterexample that is confirmed far more than challenged by Williams comments.

Williams' third point of criticism—using examples that only work against Closure under the assumption of Iterativity—is directed at the last step in Williams' reconstruction of Dretske's counterexample: the step from (3) to (2):

(3) $K(P) \ \& \ \neg K(\neg H)$

(2) $K(K(P)) \ \& \ \neg K(\neg H)$

in order to get a genuine counterexample of Closure over the entailment (1):

(1) $K(P) \rightarrow (\neg H)$.

But, for the reasons just given, this last step is never called for in order to generate a genuine counterexample to Closure. It is, in any case, surprising that Williams should interpret Dretske as tacitly relying on the principle of Iterativity, particularly in the light of Dretske's explicit denial of it:

Knowing that one knows is a form of inoculation against sceptical challenges to the quality of one's empirically conclusive reasons; one *knows* that [Diii]. Lacking such inoculation, however, one still knows. One is simply less prepared to defend the *claim* to knowledge; but inability to justify the truth of what one claims, is seldom, if ever, a refutation of the truth of what one claims, and this applies to knowledge claims as well as any other. (Dretske 1971, pp. 17-18)

The fourth of Williams' criticisms against Dretske's account is concerned with the category of cases trading on contextually generated ambiguities. However, since we already have dismissed this category of cases altogether (section 2.1), this point needs not to be pursued here.

William's overall contention is that once certain confusions are cleared away, nothing remains of Dretske's case for non-Closure. On the contrary, we can now see that the more confusions are cleared away, the more appeal is bestowed on Dretske's arguments.

2.4 Tracking

We can now proceed to a discussion of Nozick's so-called 'Tracking Theory'. As mentioned above, there is a remarkable affinity between the two theories of Dretske and Nozick.³² In its simplest version, Nozick's theory states that an epistemic agent S knows some subject matter P iff:

- (Ni) P,
- (Nii) S believes P on the basis of having implemented some P-inquiring method M,
- (Niii) $\neg P \square \rightarrow$ S does not believe P (*variation condition*),
- (Niv) $P \square \rightarrow$ S believes P³³ (*adherence condition*).

Comparing this, cursorily, with Dretske's definition:

- (Dii) S believes that P (solely) on the basis of her evidence R,
- (Diii) $\neg P \square \rightarrow$ S does not believe P,

reveals that Nozick's theory, on the formal level, only diverges from Dretske's by adding (Ni) and (Niv). However, each of these added³⁴ conditions reveals digressions from Dretske's theory which are at a deeper philosophical level. We shall discuss the two additions in turn. First, however, it is worth pointing out that

³² Which is acknowledged by Nozick. See Nozick 1981, p. 689, n. 53.

³³ *Ibid.* pp. 172-6.

³⁴ Although 'addition' is badly chosen term since Nozick developed his theory independently of Dretske. Cf. Nozick 1981, p. 689, n. 53.

there is also an interesting difference in methodology between the two philosophers. As we have seen, Dretske takes as his starting point the observation that Closure fails in cases involving contrast consequences and proceeds from that finding to a detailed analysis of knowledge. By contrast, Nozick's discussion takes its starting point in the observation that there has to be a modal aspect to knowledge, that knowledge requires a sensitivity of one's beliefs to truth over a range of relevant hypothetical scenarios, or possible worlds. He then goes on to specify which conditions have to be satisfied in order for beliefs to vary with, and adhere to, truth in nearby possible worlds (*variation* condition (Niii) and *adherence* condition (Niv), respectively). Finally, he derives the failure of Closure from the issuing 'tracking' theory for knowledge.³⁵ Dretske and Nozick thus reach the same results, but via logically converse approaches.

Why is the auxiliary (Ni) called for in Nozick's definition of knowledge? Dretske takes satisfaction of (Diii) to constitute a *factive* warrant for believing P. Hence satisfaction of (Diii) combined with a belief in P (based on R) implies the truth of

³⁵ Another noteworthy difference, and no doubt the difference that explains why Nozick's theory has enjoyed far more attention than Dretske's (despite the fact it was published ten years later) is their respective manner of presentation. Dretske covers a lot of ground in relatively few pages. Some of his ideas, although very interesting, are scarcely developed or put in sharp focus. It is from this wilderness (and richness) of ideas that one has to extract his version of a tracking theory. Nozick, on the other hand, proceeds far more pedagogically. He states his theory very clearly and expands it with sophistication and in an illuminating manner in the course of the following 70 pages. Interestingly, one can go back to Dretske's two papers and, here and there (among many other ideas) find nearly all the themes Nozick covers in his 70 pages. As an introduction to tracking theories, Nozick's text is by far the most accessible.

P. An analogue of (Ni) is therefore redundant on Dretske's account.³⁶ Nozick, on the other hand, apparently takes satisfaction of his (Niii) and (Niv) to constitute a *non-factive* warrant. That he does so has to be seen in the light of his view on counterfactual semantics. Unfortunately, and in contrast to Dretske, Nozick is remarkably brief in stating his view on counterfactual semantics. Here is what he says:

The counterfactual conditional is powerful and intuitive, not so easy to satisfy, yet not so powerful as to rule out everything as an instance of knowledge. I do not mean to endorse any particular possible-world account of subjunctives, nor am I committed to this type of account. (Nozick 1981, pp. 173-4)

He does, though, offer a detailed discussion of some points of disagreement with the classical theories of counterfactual semantics due to Lewis and Stalnaker.³⁷ In particular he states that:

The truth of antecedent and consequent is not alone sufficient for the truth of a subjunctive; Niv) says more than Ni) and Nii). - More accurately, since the truth of antecedent and consequent is not necessary for the truth of the subjunctive either, Niv) says something different from Ni) and Nii). (Nozick, 1981, p. 176)

And we may assume the same independence holds for (Niii): that (Niii) too says something different from the conjunction of its antecedent and consequent \neg (Ni) and \neg (Nii). According to Nozick's view on counterfactual semantics (Ni) is therefore not redundant.

These remarks easily drag us into *very* extensive discussions of counterfactual semantics. We shall, however, postpone a more thorough discussion of these matters until Chapter IV. For now we will restrict ourselves to two serious

³⁶Dretske 1971, p. 13. P can easily be derived from (Dii) and (Diii). Proof: assume \neg P. By (Dii) S then does not believe P. Which by *reductio ad absurdum* with (Dii) gives P.

³⁷See in particular n. 8, p.680.

difficulties connected with tracking theories' utilisation of counterfactuals. Their use presupposes a fixed and definable set of 'sufficiently close' worlds to be considered in the evaluation of (Niii) and (Niv). The notion of 'sufficiently close' worlds implies, in turn, two ideas:

World Order

Every possible world is measurable against some objective scale of 'closeness' to the actual world

And

World Border

There is some fixed hallmark of closeness somewhere along this scale of closeness: a hallmark which defines the range of *sufficiently* close worlds

Both ideas appear spurious. True, we do have some intuitive grasp of both World Order and World Border when reflecting, in a particular case, which world-scenarios are relevant for evaluation of Nozick's counterfactual and which scenarios are not. But these intuitions, regarding both World Order and World Border, vary wildly from context to context.³⁸ The picture is complicated further by the fact that conventions seem to have it that World Border in particular is sensitive, not just to context, but also to utterances of other counterfactuals within that context.³⁹

It might be objected that, in the light of the current contextual trend within epistemology, this context-sensitivity should be seen as a virtue for Nozick's counterfactual account rather than a weakness. The objection would take the form: since intuitions regarding World Order and World Border vary from context to

³⁸This point is particularly harmful for those contextualist theories, such as that of David Lewis (Lewis 1996), which presuppose a constant core component in the concept of knowledge, accounted for in counterfactual terms, which then is supplemented with further context sensitive characteristics.

³⁹See Wright 1983.

context, what counts as knowledge varies in tandem. In particular, in an epistemological context in which the usual sceptical stock-in-trade scenarios constantly are in focus, not just hazily distant, but inconveniently near, and in any case sufficiently near to be captured within the range of World Border—in such a context we possess very little, if any, knowledge. And, furthermore, this ought not disturb us too much. For once we turn to less secluded activities, and thus less secluded intuitions regarding World Order and World Border, our forfeited knowledge immediately restores itself.

On reflection, the contextual move is not a particularly attractive response to the stated difficulty. We epistemologists are supposed to be experts in matters to do with knowledge ascription. Can it really be that the very fact that we are experts, and thus have a sophisticated view on World Order and World Border, *undermines* our expertise? A worthwhile comparison may be made here with the physician who is an expert in matters to do with health and, in particular, ‘health ascription’. The physician certainly has a better developed awareness of disease scenarios in the same manner as we epistemologists have an unusually well developed awareness of hypothetical error scenarios. Does this fact in itself bring the physician to conclude that few, if any, of the people she examines are healthy?⁴⁰ Naturally not. Rather the physician’s sophisticated world view (disease-wise) enables her to undertake a more thorough investigation of a patient and thus increases the chances that her verdict is correct, i.e. that she diagnoses a disease when, and only when, that disease is present. The physician’s expertise enables her, more competently than the layman, to *uncover* whether or not her patient is healthy. If we did not consider the physician an expert in unravelling these facts, if her verdicts were as, or even less, likely to be correct as those of the layman, we

⁴⁰True, it is said that physicians, at an early stage of their career, really are rather hypochondriac and see diseases everywhere. But this, to be sure, is a stage a physician has to overcome in order to become an able practitioner.

would not readily accept that her verdict reflects how things are from a medical perspective—how things are in the medical context. And we would be much less likely to accept that she could judge a patient to be seriously ill while we also anticipate that the person’s health will be restored immediately on leaving the consulting room.⁴¹ Instead we would judge such a physician to be flagrantly incompetent; or, alternatively, that she was not concerned with health and illness at all, but engaged in an altogether different game, the vocabulary of which just happens to overlap phonetically, but not semantically, with everyday health conversations.⁴²

Given that we are happy to allow a physician expertise, why should we think differently about the epistemologist? Why should we accept such a relativity regarding epistemic facts? Given a disagreement between an epistemological expert and a layman as to whether someone really knows something, are we not inclined to claim one of the disputants must be mistaken? Granted, context-specific standards might generate unsettled trans-contextual disagreements. The sentence “Alice is tall” might, for example, be true when uttered within a pygmy tribe but false when uttered in the local basketball society. Likewise the sentence “Clinton is guilty” might be true for the Evangelical Church but false for the barrister. But the point in these cases is that the *meaning* of the relevant expression also varies across context. Indeed, this is the reason why these trans-contextual disagreements do not call for resolution. Divergence in opinion

⁴¹Compare Williams’ slogan: “Sceptic in the study—knowing on the street.” This slogan indicates the sort of instability of knowledge that issues once we cease confusing “the discovery that knowledge is impossible under the conditions of philosophical reflection with the discovery, under the conditions of philosophical reflection, that knowledge is generally impossible” (Williams 1996, p. xx).

⁴²An illuminating criticism of contextualism based on similar intuitions can be found in Schiffer 1996

between the physician and the layman, on the other hand, does call for settlement—it calls for settlement precisely because the physician commits herself to a verdict on health issues as this term is *ordinarily* understood by the layman. Now, is the epistemologist not likewise committed to attend to matters to do with knowledge as this term is understood in ordinary linguistic practice and thus committed to verdicts regarding knowledge to which the layman ought to be responsive?

This does not mean epistemology cannot be revisionistic. The suggestion is merely that such revisionistic proposals must be firmly founded on common sense intuitions about what it takes to know and thus have a built-in expectation that these revisionistic ideas also will be of relevance to those outside the philosophical society. In the contextualist response sketched above this policy is severely compromised.⁴³

Contextually inspired manners of constructing World Order and World Border are often motivated by considerations of infallibilism and scepticism: infallibilism suggests that knowledge-claims imply that all possibilities of error are precluded—admitting that one could be wrong about P is incompatible with claiming that one knows that P. Scepticism, on the other hand, points out that we never will be in a position to rule out *all* such possibilities of error—no matter how ideal epistemic conditions we find ourselves in, certain sophisticated trick scenarios will still be consistent with our total stock of information. Now contextualism suggests itself as the mediator between these two insights. According to contextualism an epistemic agent must rule out every possibility of error in order to know—but ‘every’ is understood as ‘every contextually determined, or mood determined, error scenarios’, *viz.* those error scenarios the

⁴³In fact the outcome is stronger: to admit that the epistemologist’s expertise is completely irrelevant to the layman is, in effect, finally to declare the frequently discussed ‘death’ of epistemology (Rorty (1979) and Williams (1977)).

agent happens to attend to in the present context.⁴⁴ However, one might well ask whether this diplomatic move really is called for in the first place. Does the infallibilist intuition really accord with ordinary ways of thinking about knowledge? Strictly speaking, you cannot rule out the possibility that someone, in a clever and undetectable manner, has planted a picture of Marilyn Monroe on the back of your shirt. You are attending to that possibility right now. Still, do you not know perfectly well that there is no such picture on the back of your shirt? Common-sense opinion would certainly respond affirmatively. The fact that your present epistemic situation is indistinguishable from certain error-scenarios does not in itself undermine your claim to knowledge as infallibilism will have us think.⁴⁵ What matters is that the epistemic agent is warranted *somehow* in her belief. No doubt, you could point to a multitude of potential candidates for warrants for your belief that there is no picture of Marilyn Monroe on the back of your shirt: and according to common sense, these warrants, or at least some of them, would be sufficiently strong to grant you knowledge that there is no such picture. No one would be so fussy as to call for warrants that *decisively* rule out this error-scenario.⁴⁶ But if that is correct, one major motivation for

⁴⁴ See for instance the two influential defences for contextualism by Lewis (1996) and DeRose (1995).

⁴⁵ Strong infallibilism, in its various versions, is driven by the so-called argument from universalisability. We shall not discuss this particular argument in any detail. For an exposition of it and a fair criticism, see Hale 1988 and Brueckner 1985.

⁴⁶ To be sure, this is not a disguised attempt at a thumbnail way with scepticism. One cannot make this move against any sceptical challenge: insist that although no *decisive* evidence is available against the sceptical hypothesis in case, still one is sufficiently warranted to be granted knowledge. On the contrary, what characterises the most interesting versions of scepticism is precisely their apparent demonstration that *no warrants whatsoever* are available against the central sceptical

contextualism—the attempt to mediate between infallibilist and sceptical intuitions—is misguided.

There may be other strategies available for dealing satisfactorily with the issues raised by World Order and World Border. The intention here is merely to raise these issues and to motivate the thought that one, very influential, response to them, *viz.*, the contextualist project,⁴⁷ does not offer a satisfactory solution.⁴⁸

The second divergence between Dretske's and Nozick's theories:

Nozick

- (Ni) P,
- (Nii) S believes P on the basis of having implemented some P-inquiring method M,
- (Niii) $\neg P \square \rightarrow$ S does not believe P,
- (Niv) $P \square \rightarrow$ S believes P.

Dretske

- (Dii) S believes that P (solely) on the basis of her evidence R,
- (Diii) $\neg P \square \rightarrow$ S does not believe P.

was Nozick's addition of the adherence condition (Niv). On the face of it (Niv) appears to be either redundant or fallacious. It can be understood to be either trivially true, or to be too strong to admit cases in which, intuitively, we would

hypothesis. Our remarks only concern themselves with the weaker claim that no *decisive* warrants are available for the epistemic agent

⁴⁷ To count the contextualist response to the issues raised by World Order and World Border as a solution might be considered too flattering. Contextualism merely dodges these difficulties by completely relativising the notions of World Order and World Border to the given context.

⁴⁸ Nor are these remarks an attempt on an exhaustive discussion of epistemic contextualism. Such a discussion would, for a start, have to include a thorough discussion of the *positive* arguments that drive philosophers towards the contextualist project.

conclude that S does indeed know P. This outcome is illustrated by considering Nozick's discussion of his Bank Robbery case:

The Bank Robbery case

Stance happens to get a glimpse of the escaping bank robber Jesse James' face, but only because Jesse's mask slips just as he passes Stance. (Nozick 1981, p. 193)

In this case, Stance knows that the identity of the escapee to be Jesse James, yet (Niv) is not satisfied (because, had the escapee been Jesse, but under slightly different circumstances—in particular, in circumstances where the mask did not slip—Stance would not have come to believe that the escapee was Jesse). Nozick thinks this unwelcome consequence can be handled by carefully specifying the method M which S brings in play when forming his belief. Thus, in the Bank Robbery case M should be specified as “looking and concluding it is Jesse James on the basis of seeing certain things” (*Ibid.*). This response might appear a bit fishy: if we allow for a specification of the M in play as fine meshed as ‘concluding that P on certain evidence’, will (Niv) not be rendered trivial? Will S not believe P any time that *that* method is applied? And, even worse, will (Niii) not, in the same breath, be deemed impossible to satisfy? Applying *that* method in nearby \neg P-worlds will never lead S to believe \neg P. However, Nozick supports his suggestion by referring back to an earlier discussion of methods (*Ibid.* pp. 179-88).

The crux of that discussion is that although (Niv) is designed to rule out cases where a remarkable amount of good fortune leads S to a true belief, we must be capable of distinguishing between two categories of luck: we have luck that application of M results in a veracious belief, on the one hand; and on the other, luck that application of M leads to any belief at all (*Ibid.* p. 179). The former can be illustrated by a compulsive liar who, on this one occasion, speaks truthfully; and the latter can be illustrated by a scrupulously truthful, but serene informant that only rarely speaks—but when she does always speaks the truth. Evidently, luck of the latter kind is entirely harmless. So a distinction is called for so we can

avoid the unwelcome consequences issuing from confounding these two categories of luck.

Now, how should (Niv) be formulated in order to make this distinction between harmful—knowledge compromising—and harmless luck explicit? Nozick suggests the following:

(Niv') If P were true and S were to use M to arrive at a belief whether (or not) P, then S would believe, via M, that P

The idea is thus that (Niv') only commits S to hit on P-beliefs in a limited subclass of the nearby P-worlds, namely the nearby P-worlds in which S comes to believe *something* regarding P.⁴⁹ It is acceptable that S remains ignorant as to whether Jesse James robbed the bank in nearby worlds where the mask did not slip, at least to the extent that S really remains ignorant, i.e. refrains from concluding *anything* regarding the fugitive's identity. The real content of (Niv') thus appears to be:

(Niv'') were P to be the case, S would not come to believe $\neg P$ when applying method M

In order to save (Niv) from the fallacious horn, we thus end up with a revised, and weaker, version (Niv''), which content boils down to the essential content of (Niii), viz. that S does not get things wrong regarding P in nearby possibilities. That is, it is never the case in nearby worlds that S comes to believe P when $\neg P$ (Niii), or $\neg P$ when P (Niv'). Only one further claim is required for nailing (Niv') on the redundant horn:

⁴⁹In the present discussion we will ignore cases in which more than one method is in play and S thus might form beliefs regarding P via some method different from M. Nozick (1981) offers remedies for such, for the present purpose insignificant, complications on pp. 179-86.

Symmetry

Whenever S, by applying his method, gets things wrong regarding P in nearby P-worlds he also gets things wrong in nearby \neg P-worlds, and *vice versa*.

If Symmetry holds, (Niv') and (Niii) are always satisfied in tandem and one of them is redundant.

Is Symmetry plausible? Two sorts of counterexamples come to mind:

- (i) cases where S is biased; and
- (ii) cases where the truth-value of P impinges on which method is being used

Taking (i) first. P might be some sort of unpleasant fact, for instance the fact that colleague of S disfavours S's academic skills. In that case, it might well be that S's method, be it of perceptual, memorial, testimonial or intellectual kind, is sufficiently reliable to serve him well in nearby \neg P worlds, but, due to the prejudice, leads him astray in nearby P-worlds. Or the other way round when P is a highly pleasant fact. There are two points worth noticing about these cases. Firstly, it seems, intuitively, that the question of knowledge possession in these cases hinges, not so much on whether both (Niii) and (Niv') are satisfied, but rather on the question whether S gets his beliefs right *despite* his bias; i.e. in the unpleasant P case, whether he actually comes to believe P; and in the pleasant P case, whether he would have refrained from believing P had \neg P been the case.⁵⁰

An apposite analogy may be drawn with a visually impaired person (bias prevents him from seeing clearly). It is not that he *misperceives* but rather he is only capable of perceiving momentarily, or, maybe, when the object of his perception is occasionally brought sufficiently near his eyes. When this person happens to be in one of his rare visually effective states, or when an object for

⁵⁰Just as the true criterion for a moral act, according to Kant, lies in performing it *although* unpleasant.

once is brought sufficiently near, he actually comes to know what he sees. The problem is that some other method, rationalising, say, kicks in and causes the person to create some perceptual beliefs even where none existed. Since the ‘rationalising’ procedure pulls in one determinate direction, towards pleasure, it is certain that unpleasant beliefs have the right kind of aetiology, i.e. perceptual, whereas the pleasant beliefs might have been formed either way, by perception or by ‘rationalising’. This is the reason why, in these biased cases, we find that believing P were an unpleasant P-fact to be the case, suffices for knowledge; and refraining from believing P had P not been the case, suffices for knowledge in cases where P is some pleasant fact. We have, in fact, a perfect analogy to Kantian ethics: the unpleasantness of performing some act (holding some belief) is the hallmark of its genuine moral (epistemic) value.

The second point to notice is that the biased cases are presumably most happily dealt with by identifying and separating the two methods in play—‘perception’ and ‘rationalising’. ‘Rationalising’ would not, for the reasons stated above, count as a knowledge-generating method. Whether the final output belief counts as knowledge is to be judged on the basis of weighing the two ingredient methods.⁵¹

The second category of cases in which Symmetry apparently fails, (ii), comprises cases which involve so-called ‘hinge-beliefs’; beliefs that do not issue from execution of any particular epistemic method but which, rather, must ‘stay fixed’ and thus provide the framework within which an epistemic investigation can be carried out. Nozick offers a detailed discussion of these cases (*Ibid.*, pp. 217-22). In the following we will concentrate on a particular version of one such case, the Reversed BIV case, with which Nozick originally motivates the introduction of (Niv).

The Reversed BIV case

Brainy has ended up in a BIV world. The scientist in charge of the BIVs is, though, an honest chap: he does not feed the BIVs any false information. So Brainy is fully aware of his rather awkward situation. When he, in a reflective mood, considers deep questions about life, the universe, his relation to it etc., the scientist kindly lets him recognise the harsh realities.⁵²

Does Brainy also *know* about the harsh reality that he is a BIV? His knowledge claim does satisfy (Ni), (Nii) and (Niii). He is a BIV, he believes, via some sort of intellectualisation method, that he is a BIV, and, had he been in some close \neg BIV world, he would not have believed himself to be a BIV. But, Nozick tells us, it is just too outrageous that S should be capable, given his inopportune situation, to come to know about these matters (*Ibid.* p. 173). Hence the introduction of (Niv) which is not satisfied: had S been in a nearby BIV world, with a less honest, or more thoughtful, scientist, he would have believed himself to be a more fortunate brain in a skull (BIS).

It is characteristic of these hinge beliefs that there are two methods available when inquiring about them: one 'unreliable', which, at best, circumstantially leads to true beliefs, and one reliable. But it is only one of these methods that an epistemic agent *can* apply. And which one it is, is causally determined by the truth-value of the hinge in question. For instance, in the Reversed BIV case, the agent can either maunder⁵³ or intellectualise when enquiring about his bodily status. The agent is intellectualising if he is embodied, but maundering if he is envatted. Therefore execution of either method fails to get S anywhere: he can only be said to know the upshot of the investigation if he already knows he is applying the reliable method (intellectualisation), but he only knows which

⁵¹See Nozick's discussion of priorities between multiple methods pp. 183-4

⁵²Freely improvised from Nozick (1981) pp. 176-7)

⁵³Maundering being a cognitive state phenomenologically indistinguishable from intellectualising but diverging from genuine intellectualisation by its inadequate causal origin. Hallucinating, hypnosis-victimisation, dreaming and BIVing all being examples of such a faulty causal origin.

method he is applying if he already knows the outcome of its application.⁵⁴ Other examples of hinge beliefs which share this feature are: ‘there is an external world’, ‘the world has existed for more than 5 minutes’, ‘there are other minds’, ‘I’m not hallucinating’, ‘there is more than one possible future’ and ‘nature is uniform’⁵⁵.

In the characterisation of hinge beliefs given above, it is implicitly suggested that they do not, as assumed by Nozick, constitute genuine counterexamples to Symmetry. More explicitly, it is suggested that the brain in the BIV world does not employ the same method when sorting out its own bodily status as the brain in a close BIS world. The BIV-brain maunders, whereas the BIS-brain intellectualises about these matters. And since Nozick insists that the employed method must be held constant throughout, it would seem that it is underdetermined from (Ni), (Nii) and (Niii) whether the BIV from the Reversed BIV case really knows about his sad fortune. However, in Nozick’s terminology a given method is individuated ‘from the inside’:

⁵⁴For details of this argument, see Wright’s so-called ‘PEP-argument’, Wright (1991b) p. 99. The PEP argument draws on internalistic presuppositions and might thus not appear convincing in the present context. The dialectical force of the argument is, however, irrelevant for the discussions in the present section.

⁵⁵ ‘There is more than one possible future’: The epistemic method one would like to bring in play in order to gain a warrant for this claim is conceivability. But if there only is one future one cannot be conceiving at all. All apparent reflections on what is the case in various possible futures are illusory. Hence, to find out whether there is more than one future one must be conceiving. But proper execution of that method already presupposes the existence of more than one possible future.

‘Nature is Uniform’: one finds out about this by means of induction. However, if nature is not uniform our inductive reasoning is fraud. To carry out the inductive investigation properly we must thus presuppose what we intend to show.

Usually, a method will have a final upshot in experience on which the belief is based, such as visual experience, and then (a) no method without this upshot is the same method, and (b) a method experientially the same, the same ‘from the inside’, will count as the same method. Basing our beliefs on experience, you and I and the person floating in the tank are using, for these purposes, the same method.⁵⁶

And likewise, in Nozick’s terminology, the maundering BIV and the intellectualising BIS are employing the same method.

It is doubtful whether this internalistic conception of method is consistent with Nozick’s general epistemic framework and, in particular, with his discussion of scepticism. Possibilities in which the *same* method has been exercised suddenly move inconveniently near actuality. Whenever you form some P-belief based on perception, will there not be a disturbingly close world in which you hold that P-belief based on the very same method—viz. all the worlds in which dream experience rather than perceptual experience have brought you to believe P?⁵⁷ If so, it seems that you sometimes can hold this P-belief as a result of exercising some particular method M’ and satisfy (Niii) (whenever you perceive P), and yet, at other times, hold the same P-belief based on the same M’ but without satisfying (Niii) (namely when you are merely dreaming P). That fact, in itself, does not render Nozick’s theory inconsistent. Rather it threatens to render the theory incomplete: it is no longer decidable from the mere fact that an agent veraciously has come to believe P as a result of exercising the epistemic method perceiving/dreaming whether that belief qualifies as knowledge. And, worse, it seems a very challenging task indeed to identify any auxiliary parameters that could complete the theory. Which features of a given case, other than specification

⁵⁶Nozick 1981, pp. 184-5. Nozick motivates this internalistic understanding of ‘method’ *ibid.*, pp. 232-3.

⁵⁷Dream experience and perceptual experience being ‘experientially the same’ would count as the same method for Nozick.

of the method, could possibly settle whether a perceiving/dreaming agent is tracking facts?

These considerations aside, it is not at all clear that cases involving hinge beliefs, and the considered BIV case in particular, threaten to undermine Symmetry. For a start, it is noteworthy that no specific method *is* involved in arriving at Falsification Transcendent beliefs. Nozick writes:

some statements play a central role in our continuing activities, or in our picture of the world or framework wherein we check other statements. So nested are these statements in our other beliefs and activities, and so do they nest them, that our belief or acceptance of them is (for almost all purposes) best represented apart from any particular methods [] The very centrality of the specific P means that (Niv) will be satisfied without reference to a specific method or way of believing. (*Ibid.*, p.185)

Consider a BIV who believes that he is a BIV but who, unlike our BIV from the Reversed BIV case, has come to believe this independently of the scientist's interference with his cognitive activities. For what reasons does such a BIV hold the belief that he is a BIV? The answer is presumably that he holds the belief for the same reasons that we BISs hold the belief that we are embodied, *viz.*, as the result of a non-specifiable, spontaneous urge to believe that is how things are! As mentioned in our discussion above, there is no method available, the application of which would provide some sort of non-circular evidence that we really are embodied. Furthermore, it seems that beliefs about our embodiment are not *acquired* at any particular time. We have held these beliefs all along and held them for no particular reasons.⁵⁸ Part of our cognitive acculturation apparently urges us to hold them spontaneously and non-inferentially.

It might prove beneficial to compare the BIV case with an analogous case for another belief which we hold in the same spontaneous and non-inferential manner,

⁵⁸For justification of this diagnosis see Wright (1991b) and the discussion of this paper in Appendix B to the present essay.

namely the belief that five plus seven equals twelve. Would an epistemic agent still believe this in a world where five plus seven equalled thirteen? Well, that is a weird question, for there is no such world. Or at least we cannot conceive of a world in which five plus seven equals thirteen. It is tempting, but presumably unwise, to think that even if there were such a world and an agent ended up there, she would continue believing that five plus seven equals twelve. The agent would still *conceive* of the world in a five-plus-seven-equals-twelve kind of manner. And, we may ask, what more is there to being a five-plus-seven-equals-twelve world than epistemic agents conceiving of it in exactly that manner? Hard as we may try, we simply cannot construe a mathematical counterpart to the Reversed BIV case where one of our spontaneous and non-inferential beliefs, for example the belief that five plus seven equals twelve, fails and an epistemic agent believes it to fail.

However, we can conceive of a BIV world. And if Brainy ended up there he would presumably continue believing he were embodied. Can we also conceive of a BIV world in which Brainy—newly arrived from a BIS world—believes he is a BIV? Suppose you happened to be Brainy. Could you envisage yourself realising that you are a BIV? Even in a BIV world with a very nice and honest scientist, how could the scientist possibly communicate this disturbing fact to you? Certainly not by letting you execute some embodiment-detecting procedure. There is no such procedure. And even if there were, your embodiment is not open, to you, to doubt, and you would thus never feel the slightest bit inclined to bring such a procedure in play. But cannot the scientist manipulate your brain such that you *come* to doubt you are a BIS and even *come* to believe you are a BIV? Can he not change your cognitive programming such that you no longer, somehow, spontaneously and non-inferentially, feel urged to believe you are a BIS? He presumably could since, *ex hypothesis*, he has manipulative access to the whole of your cognitive apparatus. But then, likewise, he could presumably change your cognitive programming such that you no longer, somehow, spontaneously and

non-inferentially, feel urged to believe that five plus seven equals twelve. Do you find the former manipulation any more conceivable than the latter?

2.5 Knowing More by Knowing Less

Nozick is aware of the difficulties accruing from the vagueness inherent in the two core concepts in his theory: ‘method’ and ‘counterfactual’. However, he believes this ‘leeway’ merely reflects an ordinary concept of knowledge which itself is a little rough around the edges. He writes:

There are two types of philosophical accounts of a notion. The first type classifies cases exactly as they are classified under the notion N, correctly classifying clear cases of N and clear cases of non-N, while leaving the unclear cases unclear. In terms of the conditions of the account, one can understand why the clear cases fall as they do, and why the unclear cases are unclear, either because different conditions of the account conflict, or because we can see why some of the account’s conditions do not apply clearly. The second type of account sharpens up the notion; it classifies the clear cases correctly and moreover classifies the cases whose status was left unclear by the notion itself, thereby making N more serviceable for some purposes. Accounts also can mix these features, reclassifying some previously unclear cases while leaving others to their old status. Our account of knowledge is of this last mixed sort, and the example is, I think, an unclear case which it leaves so. For it is unclear how far afield the subjunctive condition [(Niv)] is to travel in such cases. (*Ibid.*, p. 192)

Other ‘unclear cases’, however, can be clarified precisely because the ‘leeway’ allows for our intuition to be incorporated in the Nozickean evaluation procedure:

Since we have not specified a precise theory of subjunctives or specified precisely how to identify a method and tell when it is held fixed, there is some leeway in our account. It may be this leeway that enables the account to cope with these examples and other cases, by using the constituent notions loosely and intuitively. This is not an objection but a reason to think the notions can be specified more precisely to handle the cases—a condition on their specification is that they handle the cases adequately—provided the discussion of the cases did not exploit the leeway or wobble inconsistently, first leaning in one direction, then in another. (*Ibid.*, p. 193)

Such inconsistent oscillation between two incompatible directions is, however, one of the major accusations against Nozick in the extensive secondary literature. Compare, for example, Nozick's discussion of the Bank Robbery case with that in the Moody case:

The Moody case

Moody cannot distinguish between the identical twins Judy and Trudy. But as a result of a bump on the head Moody comes to believe that Judy has a mole—just as Judy, as a matter of fact, develops this differentiating feature. Intuitively Moody does not know Judy is before him even if Moody now forms a true belief to that effect (*Ibid.*, p. 191).

And, Nozick reassures us, this also is the verdict which his account offers, since (Niv) is not satisfied: "If Judy were before him, but in the very close situation of not having developed the mole, he wouldn't believe it was she." However, in the discussion of the Bank Robbery case, we saw Nozick was compelled to define the method in question as 'looking and concluding certain things' and that every method with the same upshot, from the inside, is to be counted as that same method. Hence 'By that method (applied this way) [Stance] would know in other cases also'. However, as pointed out by Forbes:

By the token of his treatment of the [Bank Robbery] example, he should rather hold that [Niv] is true [in the Moody case], since if Judy does not develop the mole, [Moody] does not see the same things, and so is not using the same method. (Forbes, 1984, p. 48)

In order to be consistent, Nozick ought thus to hold that the method used by Moody in the nearby world in which she attends to a mole-less Judy counts as a different method. Certainly, in this nearby world she is not looking and seeing a moled face—hence the method has a distinct phenomenological upshot. If, on the other hand, we grant Nozick (*Ibid.*, p. 191) that Moody is applying the *same* method in these two scenarios, must he not, on pain of inconsistency, admit that Stance in the Bank Robbery case is executing the same method no matter whether

Jesse is wearing a mask or not (and thus obtain a misleading verdict from (Niv) on the Jesse case)?⁵⁹

Apart from these well-known tensions accruing from the vagueness in Nozick's theory, there is yet another, so far ignored, but arguably, much more serious threat to the Nozickean account. And this time, a difficulty which even (inconsistently) wobbling characterisations of 'method' and 'counterfactual' cannot preclude. In order to expose this deficiency, let us first state as clearly as possible the announced denial of global Closure that the Nozickean account leads to.

Call a method which satisfies (Niii) and (Niv) a 'tracking method' and call a warrant which issues from implementing some tracking method a 'tracking warrant'. Let W^* be such a tracking warrant for some true empirical claim

(P) There is a hand before me (the reader)

And let H be a characterisation of some farfetched sceptical scenario in which even W^* loses track of P:

(H) I (the reader) am a BIV

According to Nozick, you know P when you have your tracking warrant W^* for believing P but not for believing $\neg H$. And, more schematically, whenever the counterfactual:

(5) $\neg A \square \rightarrow \neg B$

⁵⁹See Forbes 1984. For other expositions of this inconsistent oscillation, see e.g. Shope 1984 and Wright 1998c. Kripke is reported to have given a seminar devoted exclusively to counterexamples to Nozick's theory (Wright).

is not satisfied one might have a warrant for believing A and yet lack one for believing B although A entails B. In these cases, the warrant in question may keep track with matters as far as the nearby $\neg P$ -worlds are concerned, but lose track before the nearest $\neg H$ -worlds appear. Nozick, like Dretske, obviously takes this puzzling fact—denial of global Closure—to be a great advantage of the theory, since it is precisely this feature that makes it possible for the theory to deal satisfactorily with scepticism. Due to the denial of global Closure, Nozick can accept all the sceptic's premises and yet prevent the sceptic from performing any widespread sceptical damage.

This might be so, but these features of the theory certainly also have their price. One of them is the serious deficiency that one apparently can know more by knowing less. In particular one can know certain conjunctions by lacking knowledge of one of the conjuncts, i.e. one can, in certain cases, know (A&B) by lacking knowledge of A! Likewise one can know certain disjuncts by lacking knowledge of the logical weaker disjunction, i.e. know A although lacking knowledge of the logically weaker (A \vee B)! Let us see why.

We saw above that you know P but fail to know $\neg H$. Do you know the disjunction—that either P or (at least) $\neg H$? We cannot simply decide that you do on the grounds that (P $\vee\neg H$) is a logical consequence of P which you know, since that move presupposes global validity of Closure—which we just have rejected. Instead we must ask whether W^* is a tracking warrant for (P $\vee\neg H$). The proper question to ask is whether (Niv): all nearby (P $\vee\neg H$)-worlds are W^* -worlds, and whether (Niii): all nearby \neg (P $\vee\neg H$)-worlds are $\neg W^*$ -worlds. (Niv) is certainly satisfied. All nearby (P $\vee\neg H$)-worlds are $\neg H$ -worlds and W^* keeps track with (P $\vee\neg H$) in $\neg H$ -worlds. However, (Niii) fails. Not all \neg (P $\vee\neg H$)-worlds are $\neg W^*$ -worlds. The nearest \neg (P $\vee\neg H$)-worlds are ($\neg P$ &H)-worlds and hence H-worlds where M loses track of matters. Accordingly you do not know (P $\vee\neg H$), according to Nozick, although you *do* know P.

But, *very* surprisingly, it seems that you do know the *conjunction*, $(P \ \& \ \neg H)$. Here W^* does serve as a tracking warrant: (Niv) holds: all nearby $(P \ \& \ \neg H)$ -worlds are W^* -worlds. And (Niii) holds too. The nearest $\neg(P \ \& \ \neg H)$ -worlds are $(\neg P \ \& \ \neg H)$ -worlds⁶⁰ and thus still $\neg H$ -worlds. And, again, W^* keeps track with things in $\neg H$ -worlds. So, if you believe Nozick, you know $(P \ \& \ \neg H)$ although you do *not* know $\neg H$.⁶¹

Now, you—as epistemic agent with a normally functioning cognitive apparatus—might find it challenging to know P and that P implies $\neg H$ and yet refrain from knowing $\neg H$. That is one thing. Another, and even harder challenge to

⁶⁰Since the nearest $\neg P$ world is much closer than the closest $\neg\neg H$, i.e., H -world.

⁶¹That you know the conjunction of a known and an unknown belief when denial of the latter is a more ‘distant’ possibility than denial of the former, was pointed out by Kripke in unpublished lectures. Cf. Luper-Foy 1987, p. 265 n. 18. Wright brought my attention to the question of the epistemic status of the disjunction of two such beliefs. Nozick too raises the question (cf. Nozick 1981, p. 230) but immediately brushes it aside with the comment: “But this result surely carries things too far”! And we should probably agree with him on that. Nevertheless it is, as just demonstrated, an undeniable consequence of his theory! Nozick pursues under which conditions Closure under inference is valid and comes to the general conclusion that Closure over some inference is valid when the premises track the truth of the conclusion, i.e. Closure holds over an inference to Q from a known P if it is the case that: $\neg Q \ \square \rightarrow \ \neg SB(P)$. This being so since coming to believe that (Q) by inference from (P) is hooked up with believing that (P) (p. 231). Nozick adds “Knowledge, almost always, will be closed under existential generalisation [\exists -introduction]. Similar remarks apply to inferring a disjunction introduction from a disjunct” (p. 236). However, as has just been demonstrated, this is not the case for disjunctions $(A \vee B)$ where $(\neg A \ \square \rightarrow \ \neg B)$ fails.

your normally functioning cognitive apparatus is to know that $(P \& \neg H)$ and yet refrain from knowing $(P \vee \neg H)$. Try! It's very, very hard.⁶²

If you manage, have a go with the following case which is even more challenging. A thermometer functions properly within the scale from minus 20°C to 100°C but is 1 degree inaccurate for measurements below minus 20°C.⁶³ If your method (M) is reading this thermometer then, according to Nozick, you can know

(P) the present temperature is 22°C

assuming it is true. But you cannot know that

(Q) the present temperature is not minus 22°C

assuming *it* is true. Your warrant, the particular 22°C reading, tracks P—but not Q. Think of that as you like. Here is the challenge: know $(P \& Q)$ but refrain from knowing $(P \vee Q)$. According to Nozick you would succeed! According to Nozick you thus know that it is now 22°C and not minus 22°C—but you do not know the present temperature is not minus 22°C !!! Likewise you know it is now 22°C—but you fail to know that it either is 22°C or (at least) is not minus 22°C !!!⁶⁴

⁶²It is, of course, unfair to submit Nozick's account to such an internalist test since his account is, in part, an externalist theory. Nevertheless it is a serious blow to his theory if it does not pass this test, since it highlights some highly counterintuitive consequences—counterintuitive consequences that are unacceptable for internalist and externalist theories alike.

⁶³There are plenty of malfunctioning and semi-functioning thermometers around in the literature. I first came across one in Armstrong 1973.

⁶⁴For precisely the same reasons as those just exploited in the BIV/hand example.

2.6 Summary

This chapter opened with a brief exposition of Dretske's version of a modal theory of knowledge. We discussed one influential set of criticisms of it—Michael Williams'— and, in particular its rejection of Closure but concluded that these criticisms do not raise any serious threat for such modal theories. We then proceeded with a discussion of Nozick's tracking theory and noted that it differs from Dretske's in adding (Ni), P, and (Niv), $P \Box \rightarrow S$ believes P, to Dretske's theory. These two formal divergences turned out to reflect divergences on a deeper philosophical level, in particular on issues to do with counterfactual semantics and epistemic methodology, respectively. We saw that (Ni) is redundant if the classical counterfactual semantics (Lewis/Stalnaker) is accepted. Nozick explicitly denies this feature of classical counterfactual semantics, but does so without offering any definite alternative. Without such an alternative we are left with severe difficulties when attempting to substantiate the two core ideas in counterfactual semantics: World Order and World Border. The discussion of the second issue—the significance of (Niv)—let us to a discussion of the Symmetry principle: the view that the two Tracking conditionals always are either satisfied in tandem or falsified in tandem. We discussed two apparent counterexamples to Symmetry (biased cases and cases involving hinge belief) and concluded that neither demanded radical revision of Symmetry.

Apart from these general difficulties we then proceeded to a discussion of another threat against both the Dretskean and the Nozickean version of the tracking theory: the various counterexamples to the theory. The problem cases included the Bank Robbery case, the Reversed BIV case and the Moody case. Finally we focused attention on a striking consequence of the tracking theory, *viz.*, that Closure does not hold globally. We demonstrated that denial of closure of knowledge over implication implies a denial of global closure also over \vee -introduction and $\&$ -elimination. This should be considered a serious deficiency of the tracking theory as it appears highly counterintuitive. This was vividly

illustrated in for instance the Thermometer example where one (according to Dretske and Nozick) knows the conjunction that it is 22°C and not -22°C —and yet fails to know one of the conjuncts, that it is not -22°C .

Chapter III

Tracking Dispositionality

3.1 Epistemic Dispositionality and Tracking

3.2 Dispositional Concepts

3.3 Altering, Masking and Mimicking

3.4 Epistemic Mimicking

3.5 Epistemic Masking

3.6 Epistemic Altering

3.7 Internal Versus External Dispositionality

3.8 Summary

Synopsis

In this chapter we will explore an interesting analogy between, on the one hand, the various categories of counterexample to the Dretskean and Nozickian theories that accrued in Chapter II and, on the other, the various sort of counterexamples to the so-called ‘Conditional Analysis’ of dispositionality. This analogy suggests that the core idea in the Dretskean and Nozickian theories—the idea that knowledge is a matter of getting things right in a range of hypothetical scenarios—is sound despite the difficulties pointed out in Chapter II; that the real issue rather is how this insight ought to be formulated, and, in particular, that a subjunctive formulation, although providing a good approximation, comes apart from the real thing precisely in the sort of cases where the conditional analysis of dispositionality generally proves its deficiency. With this finding in place the way is paved in Chapter IV for developing an account of dispositionality which avoids these pitfalls in general—leading, in particular, to an amended version of a dispositional theory of knowledge and warranted beliefs.

We will proceed as follows: in 3.1 it is established that a dispositional capability to get things right, given the conditional analysis of dispositionality, amounts exactly to satisfaction of the tracking conditionals. 3.2 provides an account of dispositional concepts generally with the aim, in 3.3 to point out the difficulties involved in interpreting such dispositional concepts conditionally. In particular the taxonomy of the various categories of counterexamples to the conditional analysis is mapped. In 3.4, 3.5 and 3.6 analogous categories of epistemic ‘mimicking’, ‘masking’ and ‘altering’, respectively, are identified among the counterexamples to the tracking theory. Finally, in 3.7 a remarkable feature of epistemic dispositionality in particular is uncovered, *viz.*, its essential dual character: epistemic dispositional capabilities to explore the world always co-occur with certain external dispositional properties of the world explored.

3.1 Epistemic Dispositionality And Tracking

The significant insight in the Dretskean and Nozickian theories is that knowledge results from exercise of a skill to get things right; not merely in the actual world but also in a range of relevantly close worlds. In Nozick’s writings, a number of passages indicate that he might have thought of this skill as a dispositional capability. For example, he writes:

which method a person actually is using will depend on which general *disposition* to acquire beliefs [] he actually is exercising. (Nozick 1981, p. 185; my italics)

And again:

One suspects there will be some gimmick whereby whenever P is truly believed a trivial method M can be specified which satisfies conditions (Niii) and (Niv). If so, then further conditions will have to be imposed upon M, in addition to the *dispositional* condition. (Ibid., p. 684, n.21; my italics)

Although it can be questioned whether Nozick really thought about these matters in this way, it is worth exploring this idea: that an agent S knows P iff S has a veracious P-belief based on some warrant W which disposes S to get things right regarding P. It is only too natural, like Dretske and Nozick, to flesh this idea out in subjunctive terms. For, until recently, a consensus ruled in philosophical circles that dispositions should be understood in such a subjunctive fashion.⁶⁵ For example, a fragile object's disposition to break if struck traditionally has been analysed in terms of subjunctive conditionals: the object is fragile iff were it to be struck, it would shatter. Likewise poisonousness—the dispositional property to kill if ingested—has been identified with the subjunctive conditional: were a poisonous object to be ingested, it would cause death or illness. Or courageousness—the disposition to act courageously if encountering danger—as: were a dangerous situation to occur, the person would act courageously. Generally, dispositional properties have been subjected to the following conditional analysis:

Conditional Analysis (CA)

Entity O has the dispositional property to R if C iff were C to be the case, O would display R

Now, if an epistemic agent has the dispositional 'property' to get things right regarding P—i.e. (a) to believe P if P and (b) to refrain from believing P if \neg P—in circumstances where she implements some epistemological method M (M-conditions), then, according to CA, this dispositional property amounts to satisfaction of the two subjunctive conditionals:

- (a) were M-conditions to obtain and P, the agent would believe P; and
- (b) were M-conditions to obtain and \neg P, the agent would not believe P

⁶⁵But see C. Martin 1994 for an account of his solitary struggle against the CA analysis.

which are equivalent to Nozick's (Niii) and (Niv). So, given that CA offers the correct interpretation of dispositionality, the Nozickian tracking conditionals are the natural way to express precisely what such an epistemic dispositionality amounts to. However, recently CA has been subjected to severe criticism. The main reason is that there appears to be a wide range of counterexamples to this analysis.⁶⁶ These counterexamples divide smoothly into four categories: *masking*, *mimicking*, and two kinds of *altering*.⁶⁷ It is interesting to observe that the classical counterexamples to tracking theories of knowledge found in the (extensive) literature naturally divide into four analogous categories.

3.2 *Dispositional Concepts*

Before we proceed to demonstrate this taxonomical analogy, we need to do two things: First, we must say a little about the general nature of dispositional concepts; and, second, a brief elaboration is called for of the phenomena of *masking*, *mimicking*, and *altering*.

The objective is to develop a dispositional theory of knowledge. A theory which builds on the intuition underlying Nozick's tracking conception of knowledge:

⁶⁶The suggestion that tracking conditionals are symptomatic, rather than constitutive for knowledge has previously been put forth by McGinn (1984) and Forbes (1984). McGinn furthermore acknowledges some sort of conceptual tie between tracking and the "knowledge defining property". However, the candidate he finally puts forth as the defining property for knowledge, a discriminatory capability, is very different from the candidate offered in this thesis. Indeed, McGinn succumb to precisely the same mistake as the one just diagnosed in the tracking theories: not distinguishing between tracking and epistemic dispositionality (see e.g. McGinn (1984), p. 541).

⁶⁷ See e.g. Johnston (1993) and Wright (1991a).

that knowledge demands skills to get things right, not merely in the actual world, but also in a relevant range of hypothetical situations. As we have seen, Nozick himself expressed this idea in subjunctive terms as the claim that knowledge of P, in addition to P being true and some degree of belief in P, calls for covariance of truth and belief in contrafactual, but yet relevantly similar scenarios. The same core idea can, as suggested above, also be fleshed out in dispositional terms as the claim that knowledge of P is exercise of a disposition to get things right regarding P, that is, being disposed only to form true beliefs regarding P under M-conditions. In order to state this idea as clearly as possible, we need a closer examination of the nature of dispositional concepts generally.

As a starting point, we will adopt the terminology of Mark Johnston and define a concept F as dispositional just in case the following identity holds (Johnston 1993, p. 103):

Dispositional Equation

The concept F is the concept of the disposition to produce R in S under C

where R is the manifestation of the disposition, S the locus of manifestation and C the conditions of manifestation. This general formulation allows us to focus on an interesting subclass of dispositional concepts of relevance to the present discussion, which we shall term *response dispositional* concepts. A response-dispositional concept satisfies the following three conditions: (i) the manifestation R is some intrinsically mental response of some subject or group of subjects; (ii) the locus of manifestation S is some subject or group of subjects; and, finally, (iii) the conditions of manifestation C are some specified conditions appropriate for S's R-ing. Analogous to the response-dispositional concepts we have response-dispositional properties—properties properly characterised by response-dispositional concepts; response-dispositional propositions—propositions predicating response-dispositional properties of some entity; and response-

dispositional discourses—discourses essentially involving response-dispositional propositions.

Depending on purpose, we must add some further restrictions to the dispositional equivalence. Firstly, in order to arrive at a philosophically interesting notion at all, R, S and C must be *substantially* specified.⁶⁸ For the equivalence is trivially true for any F if we allow S being non-substantially specified as those subjects, whoever they are, that R under C; or if we specify C as: those conditions, whichever they are, under which S R; or, finally, if we specify R as the responses, whatever they are, which S exhibit under C. Substantial specification is more easily given for R—as for instance some judgmental response—than it is for S and C. S must be competent appraisers of F, and C must be appropriate conditions for F investigation. The inclination is to establish this by appealing to normality. S could for instance be specified as those subjects whose perceptual and other cognitive faculties function satisfactorily according to norms prescribed by statistical prevalence. And likewise C could be specified as statistically normal outdoor, out-of-shadow, daylight viewing conditions combined with a specification about S being appropriately located relative to the F subject matter and exercising the relevant cognitive skills. However, matters are not that simple.

One reason why S and C are not so easily specifiable in terms of statistical normality emerges from the following consideration. According to our definition so far, most, if not all, response-dispositional concepts applicable to external objects have a response-dispositional twin-concept applicable to cognitive entities. For instance the (arguably) response-dispositional concept of being red has the response-dispositional concept of red-*receptivity* as a cognitive counterpart: a red object O is disposed to make S judge O red under C, but also S's cognitive design is so disposed as to judge O red in response to O under C. In such dual response-

⁶⁸ In this section I follow Wright's lead from his discussion of response-dependence in Wright (1992a), pp. 108-40.

dispositional cases we can talk of a pair consisting of an *external* response-dispositional concept (e.g. the concept of being red) and a corresponding *internal* response-dispositional concept (e.g. the concept of being red-receptive).

Now, if we focus attention on the external response-dispositional concept in such a pair it is natural to expect that its correct application to some object is superveniently dependent on categorical properties of that object. In the case of colour concepts, for instance, their correct application to a particular object depends on whether this object has certain characteristic micro surface corrugations. Correct application of the internal response-dispositional twin-concept, colour-sensitivity, on the other hand, is not in this way dependent on categorical features of external objects. This can be illustrated by considering a hypothetical scenario in which the internal response-dispositional property has undergone slight changes, for instance a hypothetical scenario in which red-green colour-blindness is normal. In such a scenario we would still like to say that red things are qualitatively different from green things.

Internal response-dispositional concepts such as colour-competence are thus potentially divergent in extension from their external counterparts. So if we attempt a substantial specification of S in the Dispositional Equation in terms of standards set by statistical normality, that is, in terms of whichever colour-competence is statistically prevalent, then the two concepts featuring in the Dispositional Equation (the concept 'red' and the concept of the disposition to produce a red-response in S under C) also become potentially divergent in extension. That, in turn, means that the Dispositional Equation, at best, is contingently true. However, being a conceptual equivalence, we would expect it at the very least to be necessarily true if true at all. It appears thus highly problematic to provide a substantial specification of S in terms of statistical normality.

Furthermore since hypothetical situations where the standard C-conditions are slightly different from the present standard conditions likewise would imply changes of the internal response-dispositions (consider for instance normal C-

conditions in which electromagnetic rays get their frequencies increased due to some weird gravitational forces) the same considerations apply, *mutatis mutandis*, to a specification of C in terms of statistical normality.

Or so it seems. However, the above considerations presuppose a particular interpretation of the two concepts in the Dispositional Equation. We have:

Dispositional Equation

The concept 'red' is the concept of the disposition to produce an 'O is red' judgement by normal S under C-conditions

The Dispositional Equation must be necessarily true if true at all, i.e. (left-to-right) necessarily, red things are dispositionally judged red by normal S under C-conditions. As witnessed in the above demonstration of potential divergence in extension between internal and external response-dispositional concepts, it is not necessary that red things are dispositionally judged red by (*de dicto*) normal S under C-conditions. But we can still maintain that it is a matter of necessity that red things are dispositionally judged red by (*de re*) normal S under C-conditions, i.e. red things are necessarily dispositionally judged red in all hypothetical situations by subjects who, actually, qualify as normal (although red things are not necessarily dispositionally judged red in all hypothetical situations by those subjects qualifying as normal within the various hypothetical situations).

This ambiguity can be formalised as a scope distinction for the predicate 'normal':

$$(1) (\forall O)(\forall S) \mathbf{L} [(RO \ \& \ NS) \leftrightarrow DO\{JS(RO)\}]^{69} \quad (\text{False})$$

$$(2) (\forall O)(\forall S) (NS \rightarrow \mathbf{L} [RO \leftrightarrow DO\{JS(RO)\}]) \quad (\text{True})$$

⁶⁹Key: O = object, S = subject, C = condition, R = red, N = normal, DS(A) = S is disposed to A and JS(P) = S judge that P, L(P) = it is necessary that P.

And, again, the same considerations apply, *mutatis mutandis*, to the interpretation of ‘normal C-conditions’. A provisional formulation of the Dispositional Equation might thus be expressed formally as:

$$(3) (\forall O)(\forall S)(\forall C) \{ (NC \ \& \ NS) \rightarrow \mathbf{L} [RO \leftrightarrow DO\{JS(RO)\}] \}$$

This is, however, a rather diluted notion of necessity. In fact, the modal operator in (3) only quantifies over those hypothetical scenarios which are, in all relevant respects, similar to actuality.

And as such, it may be argued that the modal operator in (3) is rendered trivial in any case since it operates on another modal expression, namely ‘D’ in $DO\{JS(RO)\}$. It states that O is *disposed* to generate a red-sensation in S and thus bring S to judge O red. This is not a descriptive claim about what as a matter of fact will happen if that disposition is triggered but, rather, a claim about what is *likely* to happen if the disposition is triggered. To illustrate the point, consider a hypothetical scenario in which a hypnotist has manipulated you to believe the next object you attend to is green. Strictly speaking, your perceptual abilities are still functioning normally in this scenario and the surrounding conditions are still conducive for perceptual investigation. It is just that some extraordinary force has kicked in and overruled your usual colour competence. The antecedent in (3) is thus satisfied, S and C in our considered scenario are normal according to de facto standards for normality.

Now, does the existence of such a hypothetical scenario undermine the necessity-operator in (3)? The answer is no. True, the sketched scenario shows that (a red) O does not cause you to judge it red in *all* (normal) scenarios. However, the necessity operator in (3) only requires that O be *disposed* to make you judge it red in all (normal) scenarios—and that claim is perfectly consistent

with the hypothetical scenario under consideration.⁷⁰ But now we may well ask what role the necessity operator in (3) *does* play? Interpreted such as to quantify only over de facto normal scenarios (as (2) rather than as (1)) and, with reference to this limited range of scenarios, only stating that something necessarily is *likely* to happen (due to the other modal expression ‘D’) the necessity operator has been become vacuous. For how could it be merely contingently true that something is likely to happen when both ‘contingently’ and ‘likely’ quantify over the same limited range of scenarios?⁷¹

These considerations confront us with the following dilemma: either the necessity operator is interpreted as in (1) and it imposes an intolerably strong requirement (that the internal and external response-dispositional concepts are co-extensional in all possible scenarios) which renders the Dispositional Equation highly implausible. Or it is interpreted as in (2) and is rendered trivial. The natural reaction to this dilemma is to give up the claim that the Dispositional Equation is true as a matter of necessity. Accordingly we must concede that we are not really engaged in conceptual analysis at all—at least not in the traditional sense of working out (necessarily true) conceptual equivalences. Rather we must rest content with considering the Dispositional Equation as an informative elaboration on (as opposed to an exhaustive conceptual analysis of) response-dispositional concepts.

Resting content with something less than a full-blooded conceptual analysis has the further advantage that we steer free of a difficulty Johnston has articulated concerning the very idea of response-dispositionality, *viz.*, his so-called missing explanation argument. In short this argument purports to show that response-

⁷⁰We shall return to a more detailed treatment of this and germane scenarios in section 3.7.

⁷¹ More formally: due to the limited selection of possible scenarios, reflectivity, symmetry and transitivity is ensured for the accessibility relation between the scenarios. Hence, the S5 axiom **MP** → **LMP** holds (as a matter of analyticity).

dispositionality is inconsistent with the mere possibility of providing explanations of our cognitive responses—explanations, for instance of why we are inclined to judge something red. The following could be one such explanation:

Explanation: We are disposed to judge O red *because* O is red

However, Dispositional Equation, if understood as a full-blooded conceptual analysis of the concept ‘red’, ought to allow for substitution of ‘O is red’ in Explanation with ‘we are disposed to judge O red’ and thereby obtain what Johnston phrases ‘an explanatory solipsism’:

Solipsism: We are disposed to judge O red *because* we are disposed to judge O red

The upshot of the argument is that no concept to which Explanation is potentially applicable can be response-dependent.

It is interesting to notice that if one accepts the reasoning inherent in the missing explanation argument one easily slides into the opposite extreme: the thesis that all empirical concepts are physicalistic in nature:

Explanation: We are disposed to experience O as red and thus judge O red *because* O is red

But science tells us that:

Science: *Only* Physicalistic properties of O, such as its surface corrugation, can explain why we experience O as red

Hence:

Explanation’: We are disposed to experience O as red and thus judge O red *because* O has certain physicalistic properties

So if equivalent concepts—and only equivalent concepts—can be substituted in Explanation we get:

Physicalistic Equation

The concept ‘red’ is the concept of possessing certain physicalistic properties

In other words: redness *is* a physicalistic property. And generally: all empirical concepts to which Explanation is potentially applicable are physicalistic.

However, there are independent reasons for thinking that the missing explanation argument is flawed.⁷² But, as mentioned, by treating Dispositional Equation as a general characterisation of response-dispositionality, rather than as a conceptual analysis, we need not commit ourselves to a detailed discussion of these issues. For our purposes the following (non-modal) interpretation of Dispositional Equation will suffice:

$$(4) (\forall O)(\forall S)(\forall C) \{ (NC \ \& \ NS) \rightarrow [RO \leftrightarrow DO\{JS(RO)\}] \}$$

Our starting point was that a response-dispositional concept F is the concept of the disposition to produce R in S under C. We have now elaborated a little on the ingredient notions in this analysis: S, C and R. The next task is to explain what it is for O to be *disposed* to make S R under C.

3.3 *Altering, Masking and Mimicking*

This leads us to the second task announced: a brief exposition of the phenomena of *altering*, *masking* and *mimicking*. For, bracketing these phenomena, there a straightforward answer to the question what it means to be *disposed* to make S R

⁷² See for instance Menzies and Pettit (1993) and Miller (1995 and 1997).

under C, namely the answer offered by the CA analysis. The CA analysis says, to recap, that:

Entity O has the dispositional property F to make entity S exhibit response R to stimuli C iff were C to be the case, O would make S R

So what are these phenomena of mimicking, masking and altering which threaten this CA account of dispositionality?

Altering occurs when triggering of some disposition of an object causes the object to lose the disposition in question. In such altering cases the truth-values of the left-hand and right-hand side in CA come apart. The object has the dispositional property F but it would not display the characteristic manifestation R were it to be triggered since the object, due to the triggering, would *alter* and lose the dispositional property F sooner than it could bring about the characteristic manifestation. Altering is illustrated by the following example. Consider a circuit made of ‘thermo-metal’, a metal which now has its atoms alpha-bonded, and thus is conductive, but which, when subjected to a sudden change of temperature, swaps and organises its atoms in non-conductive beta-bond. This circuit is, now, conductive—disposed to conduct when subjected to electrical currents. But were this disposition to be stimulated—were the circuit to be subjected to current and, thereby, undergo a change of temperature—it would fail to conduct since this stimulation would cause it immediately to swap into its beta mode and thus alter the disposition to conduct. The very triggering of the disposition causes the disposition to disappear. Such a circuit is disposed, modulo alteration, to conduct currents.⁷³

⁷³This example is (freely improvised from) the example B. Martin (1994) initially applied in order to illustrate the phenomenon of altering. See also Shope 1978; Prior 1982; Johnston 1992, 1992a, 1993; Wright 1991a, 1992a and Lewis 1997a.

Another example of altering is provided by Johnston:

There might have been a shy but powerfully intuitive chameleon which in the dark was green but also would intuit when it was about to be put in viewing conditions and would instantaneously blush bright red as a result. So although in the dark the chameleon is green it is not true of it in the dark that were it to be viewed it would look green. It would look bright red. (1993, p. 119)

Again, the object in question, the chameleon, possesses a genuine disposition to appear green when observed. But when this disposition is about to be triggered, when some observer, call her Camilla, approaches, this dispositional property alters such that the manifestation, the green sensation, never materialises for Camilla.⁷⁴⁷⁵

Or, the other way round, a dispositional property might be brought about as a consequence of the stimulation. Johnston's chameleon, for instance, does not possess the dispositional property to appear red if observed. Yet, it would alter and gain that property if Camilla were to observe it. In this opposite version of altering no genuine dispositional property is featuring and yet the subjunctive conditional on the right-hand side of CA is satisfied.

The phenomena of *mimicking* and *masking* also provide apparent counterexamples to CA. These are also instances where right-hand side and left-hand side of CA come apart, but not due to altering under C-conditions of the object itself. Instead some interfering mechanism prevents the manifestation from displaying (masking), or some interfering mechanism brings about the

⁷⁴According to one influential view on dispositional properties, every dispositional property has a subvenient, categorical property (see e.g. Jackson 1996; Lewis 1997a and Elizabeth Prior 1985). The subvenient property for conduciveness, for example, is a certain molecular structure and the subvenient property for colour-appearance is a certain surface corrugation. If this picture is correct, *altering* of some dispositional property involves alteration of this underlying categorical property.

⁷⁵ For a more realistic version of altering chameleon properties, see Wright (1992a), p. 117.

characteristic manifestation although no genuine disposition is present (mimicking). Again an example will be helpful: consider a poisonous pill. The pill is disposed to kill anyone who ingests it. But assume Fred just has taken an antidote against this particular kind of poison. Were Fred to ingest the poisonous pill, he would not get killed. The disposition of the pill to kill if ingested is *masked* by the antidote. So, once more CA has been compromised. Or, for completeness, take the solid vase which containing dynamite powder. This vase is not fragile, it is not disposed to break if struck. Yet, were it to be struck it would break. But this disposition to break is *mimicked* by the dynamite powder.⁷⁶

The common feature of all these counterexamples to CA is that the left-hand side of CA is more 'stable' than the right-hand side. Whether an object has a certain dispositional property (left-hand side) is an issue entirely between that object, the specified C-conditions and R. But whether that object satisfies the subjunctive conditional on the right-hand side seems, furthermore, to be sensitive to O's more general settings. If there were another disposition in the neighbourhood which would change O's dispositional property (altering), or artificially either block (masking) or generate (mimicking) the R under C-conditions, this would have an impact on the truth-value of the right-hand side but not on the truth-value of the left-hand side of CA and thus generate counterexamples to CA.

3.4 Epistemic Mimicking

Now the way is paved for addressing the major task of this chapter: establishing the taxonomic analogy between, on the one hand, the classical counterexamples to Dretske's and Nozick's tracking theories of knowledge (henceforth the DN theory)

⁷⁶ For (a limited selection of) discussions on *mimicking* and *masking*, see Johnston 1991, 1992a;

Wright 1991a, 1992a; Lewis 1997a and Bird 1998.

and, on the other hand, the mimicked, masked and altering counterexamples to the CA analysis.

We will look into the three possibilities in turn. First epistemic mimicking which can be illustrated by the following example. An epistemic agent comes, very luckily, to believe a true P. Exactly as in the Gettier stories. The agent's method somehow is flawed, but on this particular occasion it leads to a true belief. Now add to the story that a necessary tie somehow has been established between the agent's flawed method for believing P and P:

The Dumb Case

Dumb comes to believe Jack's mother is a philosopher by asking a random person on the street, Jones, whether *his* mother is a philosopher. But it so happens that Jack (unbeknownst to Dumb) is Jones' brother so the method in question does as a matter of fact track facts (Brian Garrett).

The Holo Case

Holo comes to believe there is a vase in a box upon seeing a hologram of a vase being displayed on the box. The method is certainly flawed. But, add to the story that the hologram is switched on by a lever at the bottom of the box such that the hologram is always turned on when, but only when, there is a vase in the box. In that case Holo's method, watching the hologram, tracks facts (Lehrer, Paxon).

The Light Case

Flip forms a belief as to whether the light is switched on in the room next door by flipping a coin. If heads, she decides the light is on. But add to the story that her friend in the room next door cannot make up her mind whether to have the light switched on or not. In despair she has decided, unbeknownst to Flip, to let her actions be guided by Flip's coin toss (Luper-Foy). (If this is not convincing, let there be some hidden remote control mechanism within Flip's coin which switches on the light whenever the coin lands heads.)

In all these cases (and in numerous others with the same general structure) intuitions have it that the epistemic agents, Dumb, Holo and Flip, do not really

know P. Nor would we say of the epistemic agents in these cases that they are disposed to get things right under M-conditions. Their cognitive functionings and M-conditions together do nothing by way of increasing the chances that their respective beliefs are true. All the work in that respect is done, behind the back of the epistemic agent, by the wider surroundings. Yet, (Niii) and (Niv) *are* satisfied. Compare this with the situation in which an archer dispatches arrows completely randomly in all directions, but where each one of these arrows, due to some hidden steering mechanism—a strong magnet at the tip of the arrow, say—nevertheless reaches the target. This archer is not particularly competent, although he repeatedly hits the target. *He* is not disposed to hit his target. Yet, the manifestation—the arrow’s hitting the target—does materialise. But it does so due to some other disposition—the disposition of the target magnetically to attract the arrow (or the magnetic disposition of the arrow to approach the target, if you like). The archer is not competent, that is, is not disposed to hit his target, but the manifestation of this disposition, the arrow hitting the target, is mimicked by some other disposition in the neighbourhood. Similarly the epistemic agents in our cases do not know P. *They* are not disposed to get things right regarding P under M-conditions although they repeatedly are forming true beliefs about P. The manifestation of this disposition, forming true beliefs, does, admittedly, materialise. But it does so due to some mimicking mechanism in the neighbourhood.

The mimicking mechanism in the case of archery can, in principle, function in two different ways: it can be a mechanism located at the target, some sort of arrow-magnet, which secures that the target *attracts* the arrow; or it can be a mechanism located on the arrow, some build-in steering-mechanism, which secures that the arrow *tracks* the target. These two varieties of mimicking are perfectly mirrored in the Holo case and the Light case, respectively. In both cases some intervening mimicking mechanism, the hologram or the remote control, ensures that the agent veraciously, but undeservingly so, comes to believe P. But in

The Holo case it is P, the vase being in the box, which, through stimulation of the hologram, brings it about that Holo comes to believe that P. In that sense P artificially *attracts* P-beliefs. In The Light case, on the other hand, it is Flip's coming to believe that P via coin-tossing which brings it about that P. Here it is the P-belief, or, better, the belief-generating method, which artificially brings it about that P.

We can sharpen up this distinction by the following modification of the two cases: suppose the lever in the bottom of the vase-box stimulates an electrode which is implanted into the brain of Holo. When this electrode is stimulated it causes Holo to form a belief to the effect that there is a vase in the box. Likewise, whenever Flip, for whatever reasons, comes to believe the light is on in a particular room, this is registered by an implanted electrode which, in turn, stimulates a remote-control and thus switches on the light.⁷⁷ Now the analogy to the archery scenarios is complete: In the former case P (the target) artificially *attracts* a P-belief (the arrow). In the latter case it is the P-belief (the arrow) which artificially *tracks* P.

3.5 Epistemic Masking

Can we, likewise, conceive of cases of masking, cases of genuine dispositions to get things right under M-conditions but where the manifestation, getting things right would not display were M-conditions to obtain? There are such cases. And, once more, we will draw on our Zen master in order to illustrate

The Zen analogy would be something like a 'reversed' mimicking scenario, where a magnet has been implanted at the tip of the arrow, but this time the magnetic field surrounding the target has been reversed such that, far from

⁷⁷ Wouldn't it just be wonderful with such a remote-control mechanism generalised such as to register *all* our beliefs?

attracting the arrow, the target deflects the arrow just as it approaches its goal. Our competent master might, then, be as competent as you like—but yet fail to hit his target. Had he dispatched an arrow, he would have missed. It is not that his skills in any way would have failed. It is rather that some hidden forces working against his skills would then have been operative. And even a Zen master has his limitations—he would not beat the wicked powers of the magnet. Still, he is genuinely disposed to hit his target. Only this disposition is masked by the disposition of the magnet to deflect the arrow.

Likewise we could imagine a Reversed Light case in which the mimicking powers in The Light case above has been reversed: this time Flip forms a belief as to whether the light is switched on next door by implementing some standard epistemic method—observing the light switch, say. However, a friend of Flip is, unbeknownst to Flip, adjusting the lighting next door such as to contradict the outcome, whatever it is, of that method. If Flip comes to believe the light is on, her friend next door switches it off. So when Flip comes to believe P , $\neg P$ will actually be the case. And the other way round, when Flip comes to believe $\neg P$, P will actually be the case.⁷⁸ Would we not say that Flip is a competent epistemic agent, disposed to get things right regarding P regardless of the fact that the manifestation of that disposition—actually getting things right regarding P —fails to display? We would indeed. This, however, is not yet to claim that Flip knows about the light conditions next door. The manifestation of the disposition did, after all, not materialise—Flip did not come to form a true belief—and a true belief is still a necessary condition for knowledge.

Our working hypothesis is that knowledge, apart from a true belief, requires the belief in question to be the outcome of a genuine epistemic disposition to form true beliefs under M -conditions. If such a genuine disposition to form true beliefs

⁷⁸ If there is a difficulty here concerning timing we just reintroduce the electrode-stimulated remote-control—this time only on a reverse cycle.

under M-conditions is masked, the manifestation, a true belief, will be absent and since a true belief is an essential prerequisite for knowledge there will be no knowledge either. But what about a scenario where some genuine disposition to form true beliefs under M-conditions produces a true belief *despite* the presence of a masking mechanism? Take for instance the Reversed Light case. But add to the story that, on this one occasion, the friend did not pay attention and thus missed the chance to bring about darkness just as Flip came to believe the light was on. According to our working hypothesis Flip does know the light is on. She has a true belief that the light is on, and that belief is the manifestation of a genuine disposition to form true beliefs as to whether the light is on. In fact, the story now has a great deal of structural similarity with the Bank Robbery case we considered in Chapter II.⁷⁹

Still, one has some sense of unease in ascribing knowledge to Flip in the scenario just described. Arguably, part of the reason that one hesitates to ascribe knowledge to Flip in such cases is that we have become very accustomed—and maybe *too* accustomed—to the tracking Conditionals. It is thus a rather spontaneous reaction to object, with (Niii) in mind, that *were* P to be false (were the light to be off, in our case) Flip would still believe P since were P to be false the masking mechanism would have kicked in and caused Flip to hold a false belief despite the operation of a genuine disposition to form true beliefs. Hence, one is tempted to conclude, Flip does not know P. It will presumably take some re-education to free ourselves from that habitual (Niii) response. A first step towards that goal might be to reflect that the case just described in fact shares a great deal of structural similarity with the DNA case:

⁷⁹ Section 2.5.

The DNA Case

For certain DNA tests only 1 out of 100 implementations issues in a verdict for or against P. But when there finally is some verdict for or against P, the chances are 50,000 to one that the verdict is the correct one

Likewise in the Reversed Light case: *that* the method in question has not been frustrated by some masking force might be lucky. Maybe chances are as poor as one to hundred that it should not have been frustrated. But, given that it as a matter of fact has not suffered masking surgery, it is actually extremely reliable. Chances are 50,000 to 1, or even better, that it provides the correct verdict when it is not so obstructed by masking powers. The only, but, admittedly, crucial feature that distinguishes the Reversed Light cases from the DNA case is that the method involved in the latter would not issue in a warrant at all in the 99 bad cases whereas the method involved in the former would issue in a warrant, but a misleading warrant, in the 99 bad cases. In the DNA case we would, of course, admit that the agent does not know P on the 99 occasions where the method fails to issue in a warrant for or against P. But all this shows is that although S knows P on the one occasion where application of the method issues in a warrant, she only luckily knows P. Had things gone slightly differently she would not have known P. S was lucky. Granted. But she was lucky in that the circumstances allowed her method to issue in a warrant at all—not that it issued in the *correct* verdict. It is accidental that she came to form a belief regarding P. But, in believing P, it is not at all accidental that she gets things right. Why not draw the same consequences in the reversed Light case? That is, conclude that Flip knows P, is as a matter of fact disposed to get things right regarding P. But only is so accidentally?

As a second step in the re-education, it might be helpful to consider the following case:

Suppose we have built an extremely sophisticated computer, a machine from which we acquire knowledge given input for which it is programmed.

Suppose that this machine has the capacity to survey its own circuits and report malfunctioning, and that the chances of any malfunction in normal circumstances are extremely slight, but that malfunction affecting its self-surveying capacities is amongst the simplest of ways in which it can malfunction. Let P be the proposition that the computer is functioning properly. (Forbes 1984, pp. 46-7)

Now, we would certainly say that S would come to know a large variety of propositions by consulting this computer. Could not she come also to know P , that the computer is functioning properly? It is reasonable to claim she knows P , is genuinely disposed to get things right regarding P , although this disposition is prevented from manifesting by a masking disposition in the near surroundings (the disposition of the computer to mess up its own survey system under $\neg P$ -conditions).

This latter illustration of masking of an epistemic disposition presumably appears more convincing than the reversed Light case. Why? For two reasons. Firstly, it only involves masking surgery on one of the two dispositions involved in getting things right, the disposition *viz.*, not to believe P under M -conditions if $\neg P$ ⁸⁰; and secondly, because the likelihood of $\neg P$ is very low compared to the likelihood of P . However, the reversed Light case could easily be modified to suit this structure; that is, simply design the masking dispositions such as to become operative only under $\neg P$ -conditions and assume $\neg P$ scenarios are rather rare.

3.6 Epistemic Altering

Is there also an analogue to be found among the DN counterexamples to altering? The analogy would be a situation where the epistemic agent is genuinely disposed to get things right regarding P under M -conditions but also is disposed to lose that

⁸⁰That is, the adjacent disposition to believe P if P is left unmasked.

disposition *just* after the M-conditions obtain such that the manifestation of the former disposition, forming a true belief, would not materialise. That is just what characterises the following two cases:

suppose we are living in a universe in which there also exists a benevolent deity who watches over our sensory input: he has the intention to preserve this input by artificial means in the event of a cataclysm in which the material objects that actually produce it should suddenly go out of existence. Let us suppose that this cataclysm is, in fact, physically possible and that the deity has the power to carry out his intention [...] Yet, I am reluctant to say that, because of these facts we do not know that we are surrounded by material objects. (McGinn 1984, p. 531)

Or, even better:

Suppose S believes on the usual sorts of grounds that trees exist, but it is also true that if trees were not to exist he would still believe they did because the absence of trees would alter the chemical composition of the atmosphere in such a way as to interact with the chemical in his brain so as to preserve this belief in him (he gets hallucinations of trees, etc.). (*Ibid.*, p. 532)

Here too intuitions have it that the epistemic agent knows P although the tracking conditionals are not satisfied. And this time it is not a masking disposition in the near surroundings which kicks in and prevents the agent from forming a true belief in $\neg P$ when $\neg P$. The explanation why the tracking conditionals are not satisfied is rather to be found in the fact that the agent's *own* cognitive system undergoes an alteration under $\neg P$ -conditions; an alteration, furthermore, that causes the epistemic agent to lose the cognitive disposition in question.⁸¹

⁸¹In order to create a genuine altering counterexample to the DN theory, the altering will have to take place *just* after stimulation of the disposition to form true beliefs about P is stimulated, but sooner than the manifestation—the agent forming a true belief—materialises. It might be argued

We have now accomplished the major objective for this section: a demonstration that, among the counterexamples to the DN theory, there are perfectly analogies also to altering. However, we shall finish with considerations over certain other, in certain respects far more exciting, but also more controversial, manners in which an epistemic agent S's epistemic dispositions can be altered. We noted earlier that the question whether some object possesses the disposition to make S R under C-conditions supervenes on the underlying B-properties of that object and on the C-conditions. It is tempting to suggest possession of dispositional properties is a purely object-internal matter, i.e. to suggest an object possesses the property of fragility, say, if it has the underlying categorical B-properties characteristic for fragility. If this is so, there is only one way in which an object can be disposed, modulo alteration, to break if struck: namely if *the object* alters under C-conditions to lose the B-property in question. However, there is a difficulty with this suggestion; for, having these B-properties might well endow fragility on the object in its *actual* environment but fail to do so if the object is transported to a completely different environment in which physical laws are radically different from those governing the actual world. If the object were to be struck in those surroundings it might instantly turn green rather than break although, from an earthly point of view, it possesses all the necessary and sufficient B-properties for fragility. This object is fragile, disposed to break if struck. But it would lose that disposition if it were to be transported to certain far-fetched worlds. Now, imagine that the stimulus, striking the object, (somehow!) would cause the object instantaneously to be transported to such a far-fetched world. In that case the object would be disposed, modulo alteration, to break if struck. Not due to an

that the two sketched scenarios do not answer this request for subtle temporal timing. However, they could easily be adjusted to do so.

altering within the *object* under C-conditions; but due to an alteration of the C-conditions themselves under C-conditions.⁸²

Similarly for our archer: no matter how skilled he is, he would lose his disposition to hit the target if secretly transported to certain far-fetched scenarios—worlds with no gravity, say. And if there were some hidden transmogrifying mechanism which were triggered by the mere discharging of the arrow, we could say he was genuinely skilled, genuinely disposed to hit his target, but unfortunately so disposed in an altering manner: were he to dispatch an arrow he would, at that moment, dispatch it into a world void of gravity and thus miss his target. He too would be alterably disposed to hit his target due to an altering surgery which under C-conditions alters, not himself, but rather his surroundings such as to. Admittedly, our examples are now reaching the limits of permissible standards for creative case studies, but it all becomes a little more plausible when we proceed to the analogous category of altering for epistemic dispositions. This will be true in particular of those epistemic dispositions which are triggered more or less spontaneously, i.e. without any specifiable M-conditions⁸³. What we have in mind are the dispositions to believe the sort of propositions we briefly considered in Chapter II (section 2.4): the beliefs which, to conform to Wittgensteinian terminology, play the role of 'hinges' in the epistemic language-game (henceforth: H-beliefs).

It is hard to give any satisfactory account of the precise aetiology of hinge beliefs. They are just there. We cannot help but believe them. Whether or not we really are disposed to get things right regarding such H-beliefs is of course a big issue for which no simple solution is available. We shall come back to the matter in Chapter V. For now it will suffice conditionally to observe that *if* we are disposed

⁸²The idea of altering, masking and mimicking mechanisms targeting the C-conditions rather than the object itself was first brought to my mind by Crispin Wright (conversation) and is stated explicitly for masking in Wright (1998a).

⁸³ M-conditions being conditions under which the epistemic method M is being exercised.

to get things right regarding H-beliefs, this dispositional skill, just like the dispositional skill of competent archery, is one we would lose if we were secretly to be transported to far-fetched worlds *very* different in nature from the actual world every time the disposition were to be triggered. And this time the transmogrifying, far from being a sophisticated exercise in philosophical case studies, is constitutive of the CA analysis of the disposition in question: being disposed to get things right about P is not merely a matter of believing P if P. It is, furthermore, a dispositional capability to refrain from believing P if \neg P. And, as mentioned, the CA interpretation of the latter is just (Niii): were \neg P to be the case, S would not believe P. But \neg P in this case is the negation of some H-proposition and a \neg H-world is so radically distinct from the actual one that one's disposition to get things right about H here would suffer the same sad fate if one were to be transported to a \neg H world—if \neg H were to be the case—as the archer's disposition suffered when he was transported to the world void of gravity.

Let us assume for a moment that hinge beliefs can be accounted for in purely naturalistic terms. Socio-evolution has, say, effected some neuro-biological mechanism M in the brain which brings about H-beliefs. Then we can say possession of M is the relevant B-property underlying the epistemic disposition to form true beliefs about H just as molecular beta-structure is the relevant B-property underlying fragility. But we would have to say that although M bestows an agent with that disposition, and hence knowledge of H, in the actual world, the agent would lose that disposition if secretly transported to some \neg H world. In a \neg H world an agent, although possessing M, would be disposed to form false beliefs regarding H, just like an object with the B-property characteristic of fragility would have its fragile disposition transformed into a green disposition when transported to a world with radically different laws of nature. And since such a transportation is exactly what the antecedent of the CA analysis of the hinges disposition brings about, we can say an agent is genuinely disposed to get

things right about H, but *alterably* so: were $\neg H$ to be the case, S would lose that disposition and begin to form false beliefs regarding H.

One might object that such a dispositional capability to get things right regarding H in H scenarios, but *only* H scenarios, is a pretty useless one. Perhaps to be likened to possession of the disposition to act courageously in dangerous situations—but only until such a dangerous situation occurs: a dispositional capability possession of which is entirely falsification transcendent. However, on reflection, it seems more reasonable to turn the tables here and recall that every dispositional competence, including knowledge and archery, has its limitations. No matter how competent one is, there will always be some possible world transportation to which would induce failure in one's attempts. There is no obvious reason why that fact, in itself, should undermine the very possibility of coming to master some skill. One is a master when one has reached a certain level which is appreciated by one's fellow beings to deal with a sufficient variety of hypothetical scenarios for that skill to fulfil some pragmatic purpose in one's society. You become a Zen master in archery when you can bring back sufficient meat for your tribe from your hunting adventures. Or get on in the battle field. Or score sufficient points in the archery hall to impress the tournament judges or other authorities. Or whatever. Not the day you, in addition, beat the great power of the magnet or, undistracted by any kind of transmogrifying to other possible scenarios, continue to hit all targets. Likewise you know P when your dispositional competence to get things right regarding P is sufficiently high for other people to rely on your beliefs.⁸⁴ Not the day your epistemic skills have

⁸⁴But we do not accept the influential contextualist view that when doing epistemology our interest is so distinct from that of people on the street that we cease to know what they know. (as in the contextualist theories developed in Williams 1996 and Lewis 1996). Nor do we accept the view that it is the interests of the evaluator of knowledge claims, rather than those of the epistemic agent, that count (Lewis).

reached such a high level that you immediately would spot a transportation to a \neg H-world.⁸⁵ Nothing that has been said here supports the claim that we really are disposed to get things right regarding H-beliefs. We have merely observed that *if* we are, we might well be so in an altering manner.

3.7 *Internal versus External Dispositionality*

So far we have explored the structure of mimicking, masking and altering instances of epistemic dispositionality to form true beliefs and seen that the four categories of counterexamples to CA are perfectly mirrored in four analogous categories of counterexamples to the DN. However, there is a further distinction that needs to be drawn—a distinction which cuts across these categories. This distinction relates back to the discussion at the beginning of the chapter (section 3.2) of internal versus external dispositionality. It emerged in that discussion that response-dispositional properties always obtain in pairs: whenever an external response-dispositional property is instantiated, it will be accompanied by a corresponding internal response-dispositional property. It is therefore to be expected that mimicking, masking and altering will apply not only to external response-dispositionality, which we mainly have been concerned with so far, but also to internal response-dispositionality. Both internal and external mimicking, masking and altering of epistemic response-dispositions will, obviously, issue in counterexamples to tracking theories. We will end this chapter by illustrating the distinction we have in mind by cases of masking and altering, respectively, of an *internal* response-disposition.

Suppose Red has been hypnotised to believe that the next object she attends to is red. Red is not suffering from any kind of colour-blindness or similar visual

⁸⁵Craig argues convincingly for such a pragmatic foundation for epistemology in *Knowledge and the State of Nature* (1990).

disturbances. She is as a matter of fact disposed to get things right about colours. But if this disposition is triggered by attending to some green object, this will trigger the masking disposition imposed on S by the hypnotist to see the (green) object before her as red. The realisation of the stimulus thus suppresses the manifestation of S's ordinary disposition to perceive the object as green. But it does so without intervening directly with the colour of the object. Hence, *the object* retains its (external) response-dispositional quality: the disposition of the object to appear green is unmasked although Red's (genuine) disposition to see that object as green is masked and thus (temporarily) prevented from displaying its characteristic manifestation. Contrast this with a situation where a green disc has its (external) disposition to appear green masked by for instance a special light effect. Here the situation is reversed: the external disposition is masked whereas the corresponding internal disposition remains unmasked. A third option (which we shall revert to in Chapter V) is a combination of the two: a red object which, due to some external masking (such as special light-conditions) appears green to Red and, thereby, triggers an internal masking (such as a hypnotic spell) and thus causes Red (veraciously) to see the object as red.

Internal *altering* can be illustrated by an example originally introduced by Kripke: Quirky's neural network is wired up in such a way that Quirky dies immediately if he sees at a certain shade of yellow—killer-yellow (Lewis 1997a, p. 145). This implies that Quirky has an altering disposition to see killer-yellow objects as yellow: He is genuinely disposed to see such objects as yellow; but if he were to look at one, this disposition would (tragically) alter sooner than Quirky would see the object in question as yellow. But, again, this feature of Quirky's cognitive set-up—the altering of his *internal* disposition to see killer-yellow things as yellow—has no impact on the *external* disposition of the killer-yellow object to evoke yellow sensations in onlookers.

The significance of these two examples is that they remind us that the mimicking, masking and altering mechanism can influence either member of an

internal/external pair of response-dispositional properties. In the Chameleon case (section 3.3) it is the *external* response-dispositional property (green) of being disposed to cause a certain colour impression in Camilla that is altering. In the Quirky case, on the other hand, it is the *internal* response-dispositional property—the disposition to judge a given object to be a certain colour—that is altering (while the corresponding external response-dispositional property, yellow, is left unaltered). As we shall see in Chapter V it will be of the utmost importance to pay attention to this distinction between internal and external response-dispositionality when developing our theory of epistemic dispositionality.

3.8 Summary

In this chapter we set out with a definition of a concept *F* as *response-dispositional* iff it is the concept of the disposition to produce *R* in *S* under *C*. A special instance of such response-dispositional concepts is the concept of the disposition of a *P*-fact to produce a *P*-belief in an epistemic agent under *M*-conditions (conditions where some epistemic method *M* is being executed). We discussed some of the principal difficulties in giving substantial content to such a concept, in particular to give a non-vacuous account of the various ingredients: *S*, *R* and *C*. We then proceeded to elaborate upon what it is for a concept to be 'of the *disposition* to'. The classical answer, the CA analysis of dispositionality suggests that an entity *O* has the dispositional property *F* to make entity *S* exhibit response *R* to stimuli *C* iff were *C* to be the case, *O* would make *S* *R*. However, this CA analysis of dispositionality is vulnerable to four sorts of counterexamples: masking, mimicking and (two kinds of) altering.

This led us to the main task of the chapter: to bring out an isomorphism between these four categories of counterexamples to the CA analysis and four corresponding categories of counterexamples to the tracking theory. We can now

conjecture that every counterexample to the tracking theory for knowledge falls under the category of either epistemic mimicking, epistemic masking or epistemic altering.⁸⁶ This finding serves as a strong indication that the core intuition in the tracking theory—that knowledge issues from a capability to get things right not merely in the actual world but also in a range of hypothetical scenarios—may be sound, and, in particular, that the various deficiencies of the tracking theory uncovered in Chapter II may be due to the particular formulation of the tracking theory in subjunctive terms rather than to this core idea itself: the subjunctive formulation may offer a good approximation to the real thing in the same manner as the subjunctive definiens of CA, provides a good approximation to dispositionality (the definiendum in CA), although succumbing to mimicking, masking and altering counterexamples. Indeed, we saw that if knowledge is the result of rehearsing a dispositional propensity to get things right regarding P under M-conditions, then a CA analysis of that dispositional property issues precisely in the tracking account of knowledge. These two findings, put together, provide persuasive support for our suggestion put forth at the beginning of the chapter: that knowledge really is the result of executing a dispositional property to get things right under M-conditions⁸⁷ and that a satisfactory theory of knowledge

⁸⁶Is the universal quantifier justified here? We have not offered a systematic proof that shows every DN counterexample *must* fall under one of these categories. We have merely uncovered the underlying structure of epistemic mimicking, masking and altering and demonstrated that the classical counterexamples to the DN theory from the (extensive) tracking literature share these structures. Although a more systematic proof would have been preferred, our findings still serve as a solid justification for the universal quantifier.

⁸⁷ The reasoning, to recap, being the following: Assume knowledge is what one gets by execution of a dispositional capability to form true beliefs. Possession of such a dispositional capability is, if CA is correct, equivalent to the satisfaction of the two tracking conditionals. Now, CA is not, it seems, the correct interpretation of dispositionality in general; it comes close to the real thing but

either should be stated in dispositional, rather than subjunctive, terms or, alternatively, should trade on an improved version of the CA analysis, an analysis which does not succumb to mimicking, masking and altering counterexamples. The latter strategy, to search for such an improved version of the CA analysis is the task for the next chapter.

its deficiency shows up in four categories of counterexamples, namely (two sorts of) altering, masking and mimicking. So, we should not expect that the tracking conditionals capture what it is to have a dispositional capability to get things right. We should expect that the tracking conditionals come close to the real thing, but that their deficiency will show up in altering, masking and mimicking counterexamples. But, hey, that is exactly what has been established in the present chapter: we have just seen that there *are* altering, masking and mimicking counterexamples to the tracking theory.

Admitted, this finding does not serve as a knock down argument for our suggestion that knowledge really is the outcome of a dispositional capability to get things right. Still, it is a fairly solid indicator that this is so. A solid indication which will be supplemented by further arguments in Chapter V.

Chapter IV

Masking, Mimicking and Altering

- 4.1 Mimicked Masking — Subjunctive Conditionals**
- 4.2 Statistical Measure for Closeness — Establishing a New World Order**
- 4.3 Counter-Mere-Antecedent-Factuals And Counter-Entire-Factuals**
- 4.4 Mimicked Masking — Dispositional Claims**
- 4.5 A Naive Solution**
- 4.6 Lewis on Naive solutions**
- 4.7 Mimicked Masking Contexts — Quantifying In**
- 4.8 Altering Altering**
- 4.9 Summary**

Synopsis

In Chapter III it was suggested, and to some extent vindicated, that knowledge and warranted beliefs are dispositional notions: that knowledge, for instance, is a true belief which results from exercising a dispositional skill to get things right. It was also brought out that such an epistemic, dispositional skill, according to the CA analysis, can be captured in the tracking conditionals. We argued, furthermore, that the various counterexamples to the tracking theory ought to be ascribed to a general deficiency of the CA account: its incapacity to deal satisfactorily with masking, mimicking and altering phenomena. If these considerations are correct, improvement of the Tracking Theory can proceed via an improvement of the CA analysis. With this task in mind we will dedicate the present chapter to a thorough investigation of the nature of masking, mimicking and altering phenomena.

We will use masking—cases where the left-hand side of CA is satisfied but where the right-hand side falsified—as a test case for our analysis. We then

proceed as follows: in 4.1 we provide a critical discussion of masking in the light of the so-called Principle of Variable Strictness. This principle reminds us that it is a fallacy to derive a strengthened conditional such as $(P \ \& \ R) \ \Box \rightarrow Q$ from the simple conditional $P \ \Box \rightarrow Q$ unless $P \ \Box \rightarrow R$ (P, Q and R in this particular case being triggering of some disposition, absence of the characteristic manifestation, and presence of some manifestation suppressing mechanism, respectively). We observe that numerous recent arguments against CA based on alleged masking counterexamples, slip into precisely this fallacy. We argue that counterexamples to CA depend for their plausibility on a non-accidental relation between the triggering of the disposition (P) and the frustration of the expected manifestation (R).

This is opposed to standard subjunctive semantics according to which a mere *de facto* true P and *de facto* false Q suffice to falsify a P Q conditional (right-hand side of CA). In 4.2 we argue against that (counterintuitive) view and take a few preliminary steps towards an alternative subjunctive semantics which is based on statistical considerations. A final difficulty inherent in the various masking counterexamples to CA is the play they often make with a subtle ambiguity featuring in subjunctive claims. In 4.3 we disclose this ambiguity and thus disambiguate the right-hand side in CA. With the discussions in 4.2 and 4.3 we have vindicated the finding in 4.1: that genuine masking counterexamples to CA must provide for a non-accidental relation between the triggering of the disposition and the frustration of the expected manifestation.

The natural way to accomplish just that is by introducing composed objects in the masking counterexamples, that is, objects that are composed out of two parts: one part that possesses the dispositional property to be analysed by CA and another part which possesses the dispositional property to frustrate the manifestation of the first disposition whenever it (the first disposition) is triggered. But now problems beckon on the left-hand side in CA, for with introduction of composed objects it is doubtful whether the dispositional claim on the left-hand

side can be upheld. Although the simple object possesses the disposition in question, the composed object might well lack it. We appear thus to be confronted with a dilemma: either the manifestation-frustrating entity is detached from the dispositional entity. In that case there is a mere non-accidental connection between triggering and lack of manifestation and that does not suffice for falsification of the right-hand side in CA. Alternatively, the manifestation-frustrating entity is attached to the dispositional entity. But that does not suffice for satisfaction of the left-hand side in CA. In either case we do not get genuine counterexample to CA. We state (4.4) and defend (4.5) this dilemma as a general solution to the difficulties accruing from masking.

In 4.6 we compare this solution with the solution proposed by David Lewis. It appears that, whereas there is a great degree of agreement regarding the treatment of masking, there is an equally great degree of disagreement regarding the suitability of that solution as a general remedy also applicable to altering counterexamples to CA. In 4.7 we diagnose the source of that disagreement as a disagreement concerning contextual parameters in modal semantics. We furthermore demonstrate that Lewis' view on these matters is inconsistent. Finally, in 4.8, we complete our analysis by disencumbering CA from altering counterexamples after the same format as that deployed for masking.

4.1 Mimicked Masking — Subjunctive Conditionals

Until recently it was generally agreed that dispositional properties could be interpreted according to the Conditional Analysis:

Conditional Analysis (CA)

Object O has the dispositional property F to give response R to stimulus S iff
were S to obtain, O would display R

However, as we have seen, the works of Johnston and Martin on masking, mimicking and altering have severely compromised CA. It is now time for a more detailed and critical discussion of these phenomena. The first step in our investigation of the nature of masking, mimicking and altering consists in a clarifying discussion of the contention of the right-hand side of CA (the subjunctive claim that were S to obtain, O would display R). In what follows we will focus first on the phenomenon of masking. *Mutatis mutandis* all the findings here will apply also to the phenomenon of mimicking (in fact, as will become clear, masking and mimicking are just two sides of the same coin: whenever we have a case of masking regarding some dispositional property F we have, at the same time, an instance of mimicking of some other dispositional property incompatible with F. A glass merely mimicking fragility is, for example, at the same time an instance of its solidity being masked). The findings in our discussion of masking will, as we shall see, also have some bearing on our subsequent discussion of altering.

First, then, masking. The temptation we should avoid here is to be carried away by the very idea and thus too readily to categorise *any* case as masking which has it that some expected manifestation of a disposition fails to materialise. In order to avoid this seduction, we must lay down certain constraints on the construction of genuine masking cases. (And, once these constraints and their mutual relations have been explored it actually becomes doubtful whether *any* cases qualify as masking).

We begin by observing that the lack of an expected manifestation is sometimes due to lucky or, as the case might be, unlucky coincidences of events. A china vase happens, for example, to lie next to a very sensitive airbag which, if triggered will support the otherwise fragile vase just in time to protect it from shattering. If struck, the vase would not break. Or Fred happens to be immune to poison courtesy, say, of a recent intake of a substantial portion of haggis which, apart from its other well-known effects on the human body, also functions (say!) as

antidote for certain kinds of poison; and so were Fred - just now - to ingest a poisonous pill of a certain sort he would not suffer the usual poisoning symptoms. Or Sophie happens to have taken a drug with the peculiar side-effect, on rare subsequent occasions, to cause momentary colour-blindness. Were such a drug-influenced individual as Sophie to observe a green patch on one of these rare occasions, the patch would look red to her. All these cases, it may appear, provide masking counterexamples to a CA analysis of the respective dispositional properties of fragility, poisonousness and green-competence. In each case we are presented with a genuine disposition but a disposition which is prevented from displaying its expected manifestation due to the intervention of some manifestation-suppressing mechanism. We will call such intervening entities dispositional antidotes⁸⁸, the three dispositional antidotes just rehearsed being the airbag, the haggis and the drugs.

However, it is worth observing that the antidotes operating in these three cases only are operative under very specific conditions. The green patch would not appear green if it were to be observed *by that particular person influenced by that particular drug*. Likewise the poisonous pill would not cause death if ingested *by that particular haggis-eating person*; and the china vase would not break if struck *while placed on that particular shelf in the presence of that particular airbag*. The natural reaction to the three scenarios just sketched ought thus to be to point out that they are highly exceptional and consequently without any semantic significance for the evaluation of the subjunctive conditional featuring in CA. Although the poisonous pill would not kill were it to be ingested by some particular person with some particular eating habits, it might still be the case that the poisonous pill, generally, would kill were it to be ingested. We can put the

⁸⁸In adapting Alexander Bird's terminology. See Bird 1998.

point more formally by reference to the semantic principle of Variable Strictness for subjunctive conditionals:⁸⁹

Variable Strictness

$P \square \rightarrow Q$ does not imply $[(P \& R) \square \rightarrow Q]$ unless $(P \square \rightarrow R)$

The principle of Variable Strictness tells us that the outcome of subjunctive reasoning is to a large extent a function of which suppositions are being entertained. And, in particular, that the outcome of subjunctive reasoning may well differ as the suppositions we are entertaining become increasingly more specific, or, as we may express it: as our suppositions come to involve increasingly far-fetched hypothetical scenarios. Now, (restricting discussion to the example of the poisonous pill) CA merely states that a given pill is poisonous iff it would kill were it to be ingested. The alleged counterexample has it that were the pill to be ingested *and* were the person ingesting it to have been antidoted, then it would not kill. But, as we have just seen, this is perfectly consistent with the different subjunctive claim featuring in the CA analysis that were the pill to be ingested, it would kill. As the principle of Variable Strictness tells us, the two are only inconsistent if the subjunctive claim is added that had the pill been ingested, then the pill would have been ingested by someone who previously had been antidoted.

Our first finding is thus that genuine masking counterexamples to CA must involve some *non-accidental* connections between, on the one hand the disposition typically associated with the manifestation in question; and, on the other hand, the dispositional antidote suppressing the manifestation. More specifically, what is required is a non-accidental connection which ensures that the manifestation is not being prevented from displaying merely *on this particular*

⁸⁹As developed in Stalnaker 1968 and in Lewis 1973, pp. 31-5. Lewis refers a position which compromises the principle of Variable Strictness as 'the fallacy of strengthening the antecedent'.

occasion but, generally, would be so prevented from displaying were the disposition to be triggered. We can state this first finding in subjunctive terms by the claim that genuine instances of masking must satisfy both (a) and (b) below:

Genuine Masking

- (a) Were the disposition to be triggered and were some particular antidote to be present, then the manifestation would not display
- (b) Were the disposition to be triggered, that particular antidote would be present

By putting (a) and (b) together, we get the following characterisation of the sort of masker one needs in order to construct a genuine counterexample to CA:

Mask-apt Antidotes

Were the dispositional property F of some object O to be triggered, then the antidote A would be present and prevent F from displaying its characteristic manifestation R

This is an interesting characterisation of mask-apt antidotes since it has the same logical form as the right-hand side in CA: were the original dispositional property F to be triggered, then the antidote would produce R* (where R* is incompatible with R, the expected manifestation of the original disposition). Applied to the Pill case, this would give us the following characterisation of a mask-apt antidote to poisonous pills: were Fred to ingest a poisonous pill, then the antidote would prevent poisoning symptoms in Fred. Reading CA from right to left, we can thus say that the mask-apt antidote itself must possess a dispositional property, *viz.*, the dispositional property to prevent poisoning symptoms in Fred whenever Fred eats a poisonous pill.⁹⁰ And, in general:

⁹⁰Do we beg the question here—assuming the validity of CA in the course of an investigation into the very question of its validity? Not necessarily. The counterfactual in question may of course be a mere mimicking of some mask-apt antidote's disposition to prevent the original disposition from

Mask-apt Antidotes

The antidote A to the dispositional property possesses the dispositional property to prevent the characteristic F-manifestation from displaying whenever F is triggered

This characterisation of mask-apt antidotes makes it clear that *none* of the antidotes in our three alleged counterexamples to CA are mask-apt. Eating haggis might have the dispositional property to prevent poisonous symptoms in Fred whenever Fred eats a poisonous pill *and* eats the haggis. But the haggis, since detached from both Fred and the poisonous pill, does not have the general dispositional property to prevent poisonous symptoms in Fred whenever Fred eats a poisonous pill. Nor can the drug in our Drug case be attributed the dispositional property to make a green patch appear red whenever Sophie looks at it. Granted, the drug becomes so disposed once Sophie has taken it. But, until then, spatially, temporally and causally independent of Sophie as the drug is, it exerts no influence over Sophie—least of all is it disposed to impair her visual competence whenever she looks at a green patch.

It might be objected here that our requirement that there be a non-accidental relation between the dispositional property and its masking antidote is far too demanding. Why require that a mask-apt antidote does not merely *happen* to suppress the manifestation in question but furthermore is *disposed* to suppress the manifestation in question? It could, for example, be insisted that our three scenarios, far from being very rare exceptions without any significance for the evaluation of the subjunctive conditional in CA, *ex hypothesi*, merely reflect how the world actually is. And if the world is, *ex hypothesi*, as described in these three

displaying its manifestation and thus fail to mandate a right-to-left reading of CA. However, as applied in a mask-apt antidote, this is a distinction without a difference. Like a double-negation sentence, a mimicked masked disposition collapses into the real thing. This point is carefully argued in Johnston 1992a, pp. 16-17.

scenarios, then, since these are thus the closest of all hypothetical scenarios, they certainly ought to be taken into account when evaluating the corresponding CA subjunctive conditional.

However, hypothesising in this manner is, in effect, tacitly to assume a *de re* reading of the expressions occurring in CA. It is, in fact, to claim that for example this *particular subject*, as a matter of fact would not suffer poisonous symptoms if he were to ingest a poisonous pill. And such a *de re* reading does not comply with ordinary ways of speaking and thinking about dispositions. What is at stake when we speak and think in dispositional terms is whether the poisonous pill is disposed to cause poisonous symptoms in the subject, whoever it may be, whoever actually ingests it. From a dispositional point of view, it is less interesting to consider what would happen if *that* particular person with *that* particular background, were to ingest it. If we thus swap perspective from the *pill's* dispositional properties and instead focus on a particular *subject's* reactions to that property we do, at the same time, swap the focus from the dispositional poisonous property of the pill to the corresponding dispositional poison-sensitive property of the subject. Granted, a haggis-antidoted individual does not possess this dispositional poison-sensitive property. But that has no consequences whatsoever for the pill's dispositional poisonous property. Neither does it have any consequences for the subjunctive fact that, were someone, randomly chosen, to ingest the pill, he would come to grief.

The point might of course be pressed that our three scenarios, far from being a product of hypothetical stipulation, are simply true descriptions of some actual states of affairs. On some rare occasions it so happens that the person, whoever it may be, who ingests a poisonous pill does not suffer the usual symptoms of poisoning for the simple reason that he just has eaten some haggis. Not in some possible world but in the actual world. And, again, being an actual case and thus the closest of all possible scenarios, it should be taken into account in the evaluation of the CA subjunctive conditional. But such an objection is also misguided. To admit that ingestion of a poisonous pill on some rare occasions

fails to cause the expected symptoms is not yet to grant that consumption of the pill would not effect the expected symptoms. Rather it serves as a reminder that subjunctive implication is a weaker logical relation than strict, or necessary, implication. Even a Zen master in archery would, on some rare occasion, miss his target. That fact, on its own, does not undermine a subjunctive claim to the effect that he would hit his target were he to dispatch an arrow. In fact, to point out that the Zen master would miss the mark *on some rare occasions* is just an alternative way of stating that he, generally, would hit the target.

4.2 Statistical Measure for Closeness — Establishing a New World Order

In the foregoing we have drawn rather heavily on the notion of ‘accidental circumstances’⁹¹. But what does it mean that the obtaining of some state of affairs, such as the operation of a particular antidote, is accidentally rather than, as we have formulated it, dispositionally determined? So far the implicit assumption has been that subjunctive semantics provides the resources to settle this intricate question. Indeed, the contention of the principle of Variable Strictness, as we have set things up, is that specific mention of the accidental states of affairs is tantamount to entertaining hypothetical scenarios which are more ‘distant’ according to standard subjunctive semantics, whereas specific mention of non-accidental states of affairs involves hypothetical scenarios which standard

⁹¹And, to be sure, ‘accidental’ is intended to mean accidental—i.e. used in its ordinary sense—as opposed to its more technical meaning in recent work within philosophy of science where ‘accidental’ is taken to mean ‘not law-governed’. In the present context it thus is not accidental at all that, for example, a gold ingot is less than 8 meters long although this by no means is a law-governed fact (no laws of nature suggests that gold ingots cannot be 8 metres long).

If we use ‘accidental’ in a technical sense at all it would rather be in diametrically opposed sense as, precisely, ‘law-governed’—governed, namely, by *statistical* laws to be accidental.

subjunctive semantics considers ‘nearby’. However, it may be challenged whether it is legitimate to place such a heavy theoretical burden on subjunctive semantics, particularly when reflecting how murky and deceptive an area subjunctive semantics itself sometimes appears to be. Sometimes it might even strike one as an open question whether standard subjunctive semantics provides for assignments of objective truth-values at all. Suppose for example that someone while discussing our haggis case stubbornly insisted that Fred, had he ingested the poisonous pill, also would have eaten some antidoting haggis. By which means can we convince such a person that this is a false assumption? Is there any rational way to settle such a subjunctive disagreement? The answer of course hinges on the correct answer to the related question whether there is any objective measure for a hypothetical situation’s ‘closeness to’ or ‘likelihood in relation to’ the actual world. The assumption so far has been, and shall continue to be, that there are such objective measures.

However, even granted that standard subjunctive semantics can provide for objective truth-values for subjunctive claims—truth-values, that is, determined merely from the content of the propositions featuring in the subjunctive conditional and objective facts about the world—it may still appear rather presumptuous to claim that subjunctive semantics can carry the burden of demarcating between accidental and non-accidental states of affairs. One (rather dubious) way of motivating this presumptuous claim would be the bold suggestion that accidental states of affairs are simply more ‘distant’ than non-accidental states of affairs. Hence, since it is accidental that Fred had a portion of haggis shortly after he ingested the poisonous pill, then given that he did so, we are entertaining a more ‘distant’ scenario, and indeed one which fails to mandate the subjunctive claim that had Fred eaten the pill he would have eaten a portion of haggis. A bold suggestion along these lines is stillborn for the simple reason already repeatedly rehearsed: that such accidental coincidences, although accidental, nevertheless *do* obtain and, when they do, do so disturbingly ‘nearby’, *viz.*, in the actual world.

But a slightly more sophisticated version of the same train of thought has it that the issue of ‘closeness’ featuring in subjunctive semantics can be settled by statistical means; and, accordingly, that accidental states of affairs, *although actual*, ought to be considered ‘distant’ and thus, again, without semantic significance for our subjunctive reasoning if they are statistically exceptional. According to this line, an accidental coincidence of events, such as Fred ingesting both a poisonous pill and a portion of haggis, is considered statistically rare. In other words, focusing on all the instances of Fred having ingested a poisonous pill, the proportion of those instances in which he also has some haggis is relatively small—‘relatively’ of course to be specified in some substantial manner. The difficulty for such an approach is, though, not simply the play it makes with the notion of ‘relatively small proportion’, but rather the play made with the notion of ‘all’. Which are *all* the instances in which Fred ingests a poisonous pill? To take ‘all’ to be all *actual* cases will not do the trick. Suppose, plausibly in this particular context, that Fred only ingests one lethal dose of poison. In that case the proportion of *all* the instances of Fred ingesting poison in which he also eats some haggis is either 0 or 100 per cent! And neither would be a particularly reliable indicator as to whether or not the co-ingestion of haggis was accidental or not. So, even granted that the semantic core notion of ‘closeness’ can be put under some control by statistical means, the challenge still remains to work out the statistical frequency of some particular event pattern.⁹²

Different cases allow for different degrees of alliance of accidentalness with statistics. Suppose Fred visits a casino just on one single occasion in his entire life. Suppose, furthermore, that he plays the roulette wheel ten times on that particular occasion. And, to give the story a happy ending (for a change), suppose, finally, that he wins all ten bets. In that case he wins his bet in a very high proportion, *viz.*, 100 per cent of all the instances in which he bets. Nevertheless we would like to

⁹²Thank God we can leave this part of the theoretical burden for the statisticians.

say that his winning was an extraordinarily accidental coincidence. However, this case, as opposed to the haggis case, allows us to ascend to second-order considerations and insist that Fred's winning really was extraordinary; and beef up this claim by specifying that an accidental event is merely accidentally reproduced if it constitutes a statistically rare *frequency* of correlated events—in our particular case that among all actual instances of an individual betting ten times successively on a roulette, the proportion in which that individual also *wins* ten times successively, is relatively small.

Unfortunately accidents sometimes ascend to second-order in tandem. It might so happen that all instances of people betting ten times successively on a roulette-wheel are instances where these people, miraculously, win. I beg your pardon? Very unlikely? Of course it is! Extraordinarily accidental. But to concede this much is only to ascend yet one level higher. To block a potential infinite regress we must, at some level, insist not merely on unlikelihood but on *falsehood*. In the roulette case we do have this opportunity: It is not (merely) unlikely, but an *established* falsehood, that all ten-frequency gamblers win. However, not all accidental states of affairs allow for such higher-order statistical considerations. Accidental coincidences between gambling and winning do. Other accidental coincidences such as that between poison and haggis consumption⁹³ do not—or at least, do not do so in any obvious manner.

⁹³Not to mention co-occurrences of Jones being confronted with a dangerous situation where he responds courageously to that situation. We should not, though, get too discouraged by the poor epistemological prospects involved in disclosing the exact statistical details in all such co-occurrences. At present we are concerned with the metaphysical aspect of the question of what a counterfactual fact *is*. Eventually, in order to put the answer to that question to use in our broader context, it will be essential also to provide means for answering the related epistemological question of how it is determined *whether* such counterfactual state of affairs obtain. However, as we will see, the scope of co-occurrences we will be concerned with here—more specifically co-

A further worry which ought to be considered is this: even for the most exceptional sequence of incidences, do we not still want to say that in *some* possible scenario just that sequence is instantiated? That there is some possible scenario, for example, where *all* coin tosses hitherto have issued in heads? In such a hypothetical scenario, the statistical account of accidentalness would be misleading. But, in a sense *any* possible scenario, including the one instantiated in the actual world, has its own unique, and highly accidental, sequence of complete coin-tosses. In fact, each of these complete sequences of coin tosses is unique and, in any case, as likely, or unlikely, as any other sequence (including the sequence consisting of nothing but heads).⁹⁴ So, by way of analogy, the statistical account of accidentalness is misleading in the actual world too. And it does not count, at this point, to object that the actual complete series of coin tosses is not an instance of an accidental coincidence at all since there are no two constant entities—a coin toss and a fixed outcome—that could be so coincidentally related. There are two such constant entities. There always are! For the outcome of the coin tosses can be held fixed in disturbingly many different ways. It can be held fixed as heads. Or it can be held fixed as tails. Or it can be held fixed as alternately heads and tails. Or alternately twice heads and twice tails. In fact it can be held fixed relative to any function you like. And for each possible sequence of outcomes there will be (at least) one function relative to which that particular distribution, magically as it may appear, manifests fixedness. It can, obviously, be objected once more that it is extraordinarily accidental, maybe even suspiciously accidental, that it is precisely that one particular function relative to which particular outcome of the coin tosses stay fixed, that it is *that particular* function we have chosen to use as our measure

occurrences of implementation of some epistemological method with the forming of a true belief—will be limited to cases where we as a matter of fact have a pretty good intuitive grasp of whether the counterfactual claim in question is warranted or not.

⁹⁴ Thanks to Stewart Shapiro for pointing this out.

for accidentalness. But this is just yet another ascent to a higher level—only this time a second-order ascent along another dimension. And we can, if we wish, run the same dialectic all over again.

As is only too obvious, these are merely the first few moves in a highly complex dialectic on these matters.⁹⁵ Mention of them here is not an attempt to offer a knock-down argument either for or against the thesis that a substantial notion of accidentalness can as a matter of fact be developed on a statistical basis. Rather these considerations serve to indicate the complexity involved in developing such an argument. Our main concern is how—granted that such a statistically based concept of accidentalness can be developed—subjunctive semantics should deal with accidental states of affairs and, in particular, how it should deal with accidental occurrences of antidotes. The suggestion above was that subjunctive semantics should be rendered immune to such accidents and that it should be rendered so immune by adapting a statistical interpretation of ‘closeness’. This idea is, in effect, tantamount to adding the following ‘accidentalness clause’ to CA:

Accidentalness Clause

Object O has the dispositional property F to give response R to stimulus S iff were S to obtain, O would display R, or O would not display R—but only because some specifiable accidental circumstances prevented it from doing so⁹⁶

Nozick has remarked that it is an (unacceptable) consequence of standard subjunctive semantics that the mere truth of P and Q confers truth on the P–Q

⁹⁵For a few more moves consult McCall 1994 and Gundersen 1997. In Gundersen 2000b, Goodman’s Gruesome Modal Fallacy, I discuss the legitimacy of holding one particular function fixed above some, from a strictly logical point of view, equally deserving alternative.

⁹⁶This formulation is due to Crispin Wright.

subjunctive conditional.⁹⁷ Nozick's worry is, precisely, that standard subjunctive semantics misleads us to believe that P and Q are subjunctively related in cases where their co-obtaining is a mere accident. By way of amendment, he suggests we should beef up the necessary conditions for satisfaction of P-Q subjunctive conditionals: his suggestion is that Q has to be true, not merely in the 'closest' P-world, as standard subjunctive semantics has it, but also in the closest 'P-band' of worlds where the closest P-band of worlds, roughly, are those P-worlds closer to actuality than any \neg P-world. Nozick illustrates his point by the so-called 'double slit' experiment. An electron is fired, let this be event P, and passes through the right-hand slit, Q. And yet, he rightly points out, we ought not to say P subjunctively implies Q such as standard subjunctive semantics will have it. P only subjunctively implies Q if Q also holds in all the other P-worlds closer to the actual world than the closest \neg P world. Nozick's point gains even more weight in the case where our Zen master dispatches an arrow but, for the first time in his glamorous Zen career, misses his mark, due, say, to a sudden strong side wind. That isolated incidence does not at all warrant a subjunctive claim to the effect that were the Zen master to dispatch an arrow, he would miss. In the Zen case it even appears plausible to go one step further than Nozick does, *viz.*, to claim that dispatching an arrow subjunctively implies hitting the target *despite* an actual instance of dispatching and missing.

Are we taking things one step further with this latter suggestion? According to Lewis we are. He writes:

I have claimed that the subjunctive conditionals with true antecedent and false consequent are false, and that those with true antecedent and true consequent are true. I am fairly sure of both claims, but surer of the first. (1973, pp. 28-9)

⁹⁷Nozick 1981, pp. 680-1, n. 8. See also his footnote on p. 174.

This is so, Lewis comments, since the former is a consequence of the assumption that

(1) No world is more similar to the actual world than the actual world itself

which he takes to be perfectly safe, whereas the latter is a consequence of the assumption that

(2) No possible world is even as similar to the actual world as the actual world is to itself

which he takes to be less perfectly safe (although, of course, still fairly safe).⁹⁸

Both these claims concern the idea of World Order discussed in Chapter II—the idea that every possible world is measurable against some objective scale of ‘closeness’ to the actual world.

Our suggested statistical foundation for closeness relations implies a rejection of both (1) and (2). According to our suggestion a hypothetical situation can be even closer to the actual world than the actual world herself! Crazy? Not necessarily. When we talk about people, towns and nations it makes perfectly good sense to say “X is not at all herself today” implying that there is a more genuine version of X who, today, is a mere non-instantiated counterpart of her. Does it not make equally good sense to make such a claim about the entire world?⁹⁹ Is it not, indeed, a mere consequence of someone in particular not quite being itself that the world as such is not quite herself? We can say that it is not at all like Billy to beat

⁹⁸Presumably he has in mind here the same set of worlds as those Nozick intends to capture with his ‘P-band’ notion.

⁹⁹Cf. Hegel.

up his wife¹⁰⁰ and thereby indicate that he was not quite himself when he did so and, indeed, that the non-instantiated counterpart of Billy in some possible world who did not beat up his wife is much more similar to the actual Billy. Can we not likewise say that it is not at all like the world to let Fred have ten subsequent wins on the roulette-wheel—that it would have been much more like it to let Fred lose some of the times he bet and, indeed, that a non-instantiated counterpart of it who did not let Fred win continuously is behaving much more in accordance with its true self, is more similar to the actual world? What is significant for the evaluation of P–Q subjunctive conditionals, on the view we are developing, is what is going on in the P-band of worlds or, (even) more metaphorically, what is going on in the world as it is when it is itself most. What happens in the actual world when it is not at all like itself (when some extraordinary circumstances kick in) is irrelevant.¹⁰¹

Although the further move, challenging (1), appears *prima facie* so much more controversial than Nozick's initial move, challenging (2), the further step is, really, a mere logical consequence of the former. For once it is admitted that there can be irrelevant dissimilarities between worlds such that irrelevantly divergent possible worlds can be as close to the actual world as the actual world herself, the semantic fulcrum has already been transferred from the notion of 'similarity' to the notion of 'relevance'. Giving up (2) is giving up the idea that World Order is a function exclusively of similarity. It is, in effect, to turn World Order into a

¹⁰⁰It is just that, very unfortunately for him, and for his wife, he bumped into some statistically extremely divergent situation that a few extra electrons fired in that particular corner in his brain. Billy's brain is so designed that in any scenario governed by statistical normality, no such thing happens in that corner. It is tough luck indeed that this scenario cropped up—and its upshot in violent behaviour is one we do not associate at all with Billy's (real) character.

¹⁰¹Neither metaphor 'P-band' nor the metaphor 'being one self' is, though, called for. We can state in literal terms what we are talking about, *viz.*: statistical normality.

function of both similarity *and* relevance. For once the door has been opened in this way to irrelevant dissimilarities, how could we possibly prevent irrelevant similarities from slipping in also? If statistical frequency is the correct measure for relevance when judging which dissimilarities are irrelevant—as Nozick’s double-slit example indeed suggests they should be—ought it not likewise be the correct measure for the relevance of similarities? If so, the natural consequence is to judge statistically exceptional similarities as more distant than statistically compliant dissimilarities when defining World Order.

The obvious advantage of the statistically inspired view of subjunctive conditionals we are developing is its resources to deal satisfactorily with the clause in Accidental Clause:

Accidental Clause

Object O has the dispositional property F to give response R to stimulus S iff were S to obtain, O would display R, or O would not display R—but only because some specifiable accidental circumstances prevented it from doing so

We are now in a position to specify ‘accidental circumstances’ more substantially than simply ‘those circumstances where an expected dispositional manifestation fails to display’. We can offer such a substantial account in terms of independent, statistical measures for accidentalness. Another advantage of the statistically inspired view of subjunctive conditionals is that it offers a solid theoretical foundation for our intuitions underlying the Accidentalness Clause. Given the statistical basis for subjunctive semantics, stimulation of some disposition never subjunctively implies the obtaining of some extraordinarily accidental incidence. With that finding in place, CA and Accidentalness Clause become equivalent.¹⁰²

¹⁰² The right-hand side in Accidental Clause says that were stimulation S to obtain, then response R would obtain unless some accidental state of affair A would obtain. But our statistically inspired semantics informs us that were S to obtain, A would not obtain. Hence, given Accidental Clause,

With the introduction of a statistically based measure for the closeness relation in subjunctive semantics, we have taken a significant step towards a fact-independent understanding of CA. A masker cannot simply adhere to a disposition that, as a matter of fact, has been stimulated but failed to display its characteristic manifestation. A masker must, additionally, establish that the disposition would have failed to display its manifestation were it to be triggered. And this she cannot any longer establish from just *one* actual manifestation failure. However, the masker might, at this point, decide that she could just as well take some advantage of all this talk about accidentalness and let it play into her own hand: she could thus point out that had things gone slightly differently, had some disposition which as a matter of fact suffered manifestation failure due to the operation of some dispositional antidote, had this disposition been triggered in some slightly different way, chances are that the dispositional antidote would still have prevented the manifestation from displaying. For, the reasoning would be, minor modification of the actual course of events would have had no impact on the functioning of the antidote in question. Given that the haggis prevented Fred from dying, it would still have done so minor changes in the actual course of events notwithstanding. Hence, had Fred eaten the poisonous pill, the haggis would still have done its job, and Fred would not have suffered any poisonous symptoms. The subjunctive link requested in Mask-apt Antidote between triggering of the disposition and the intervention of the antidote appears thus to have been established and, furthermore, to have been so established on the basis merely on an actual case. To prevent this move we need to take yet another step towards a fact-independent reading of CA.

were S to obtain, R *would* obtain. But that is just the right-hand side in CA. The right-hand sides in CA and the Accidental Clause are thus equivalent.

4.3 Counter-Mere-Antecedent-Factuals And Counter-Entire-Factuals

This further step is a reminder of a crucial, but often ignored counterfactual distinction: the distinction between counter-merely-antecedent-factuals and counter-entire-factuals. Before we state this distinction in detail, let us stress what the two have in common: counterfactual investigations of both kinds never concern themselves with what the case *is*. Once it is established that such and such *is* the case, the only interesting question left is whether such and such also *implies* something else. Inquiring counterfactually, on the other hand, is inquiring *counter* factually. It is inquiring what would happen, not if such and such is the case, but, exactly, if thus and so, contrary to facts, *were* the case. We are interested in learning what would happen *were* the Zen master to dispatch an arrow—not what is happening with the arrow which he *has* just dispatched. Likewise our counterfactual interest in Fred's eating habits is not at all directed towards his actual diet. Rather we are interested in what would have happened, had he consumed something he in fact did not, *viz.*, a poisonous pill: would he, *then*, also have had a haggis?

Reasoning counterfactually, we are sometimes investigating whether some 'counterfact' would bring such a 'disturbance' into the course of events¹⁰³ so as to prevent some *fact* from occurring. In these cases we pick a true counterfactual consequent—Fred having some haggis, say—and think counterfactually only as regards the antecedent—i.e. we say *were* Fred to ingest poison—in order to investigate whether such a counter-assumption has any impact on the actually true consequent. Fred actually ate some haggis today. Would he still have done so if he had ingested poison? Sometimes the point is thus to reason counter-merely-antecedent-factually.

¹⁰³How much counterfactuals 'disturb' course of events both forward and backwards in time is elaborated upon in Lewis 1979.

At other times, however, we are simply investigating whether there is a subjunctive relation between antecedent and consequent. In that case we are inquiring counter-entire-factually. These examples will serve to make the distinction we have in mind clear:

Counter-entire-factuals

- (A) Fred would have won the lottery if he had had an extra bowl of Cornflakes for breakfast
- (B) Willie would have broken his leg if a black cat had crossed the road
- (C) Sophie would have finished the marathon if she had worn Nike shoes

And their corresponding:

Counter-mere-antecedent-factuals

- (A') Fred would still have won the lottery if he had had an extra bowl of Cornflakes for breakfast
- (B') Willie would still have broken his leg if a black cat had crossed the road
- (C') Sophie would still have finished the marathon if she had worn Nike shoes

There is a clear difference in content between counter-entire-factuals and counter-mere-antecedent-factuals. It is tempting, but incorrect, to think that it is the 'still' in the consequent, or alternatively an 'even' in the antecedent that indicates whether a counterfactual is intended as a counter-entire-factual or as a counter-mere-antecedent-factual. In certain cases such as

- (3) John would have remained calm if she had told him

a counterfactual can be ambiguous between the two readings despite the absence of a 'still' or 'even' indicator. In certain cases it is thus only context that reveals which of the two senses is intended. Likewise it is tempting, but incorrect, to think that counter-mere-antecedent-factuals are true *because* the corresponding counter-

entire-factual is false. The relation between the two is much subtler than that. But our motivation for paying attention to the distinction between counter-mere-antecedent-factuals and counter-entire-factuals is not to explore the relation between them. We only mention it here in order to nail down the point that the subjunctive conditionals featuring in CA are counter-entire-factuals.

In constructing a masking counterexample to CA it does not suffice to point out that, in the actual world, it sometimes happens that O has the dispositional property to R if S and yet would not R if S due to the antidote A. The thought in such cases is that the disposition has been stimulated, but that the antidote has been operating and that the characteristic manifestation accordingly has not displayed. Hence were the disposition to be triggered (in a slightly different manner) the intervening of the antidote would still be operating and the manifestation would thus still have been absent. This is so since the disturbance in the actual course of events issued by a slightly different stimulation of the disposition would not have any influence on the working of the antidote, and hence, would still have prevented the manifestation from displaying. This is sound reasoning only when dealing with counter-mere-antecedent-factuals.¹⁰⁴

Would lethal doses of poison cause death if ingested? This obviously is a counter-entire-factual inquiry. In responding to it, the worry that sometimes, for some people, springs to mind is this: does the answer not depend on whether the victim already had been antidoted? The principle of Variable Strictness tells us that it *does* depend on whether the victim was antidoted (if the victim were to

¹⁰⁴In these discussions we have bracketed a third category of subjunctive conditionals, *viz.*, those with actually true antecedent and actually true consequent since they are not significant for our particular purpose. English is only suitable for expression of counter-entire-factuals. In Latin e.g. all three categories and, in particular, their slightly distinct connotations are much more easily articulated (so scholars have informed me). Linguistic awkwardness, though, is no excuse for omitting a discussion of the truth-conditions for counter-mere-antecedent-factuals.

ingest poison *and* the victim had been antidoted, then the victim would not die)—but *only* is so depending upon whether the victim has been antidoted to the extent that the victim's ingesting the poison counterfactually implies his being antidoted (it is still true that the victim would die if he ingested poison, *unless* it is the case that were the victim to ingest the poison, he would be antidoted). At this point it is an incorrect move to insist that the victim, if as a matter of fact antidoted, (still) would have been anti-doted, had he ingested the poison. It is an incorrect move since it is counter-mere-antecedent-factual reasoning in the course of a counter-entire-factual investigation. The only consistent continuation of that investigation is counter-entire-factually to reflect whether the victim's ingesting the poison counter-entire-factually implies the victim's ingesting haggis. Empirical details of the actual case are irrelevant here.¹⁰⁵

4.4 Mimicked Masking — Dispositional Claims

With these findings to hand we can now proceed to the more urgent question of whether there are any genuine instances of masking at all? It is natural to suggest that although no genuine masking is involved in the cases discussed so far, these cases could very easily be modified so as to be transformed into genuine cases of masking. All it takes is a little mereological surgery.¹⁰⁶ We simply compose a

¹⁰⁵This does not mean the actual world is irrelevant for evaluation of counterfactual claims altogether. We attribute less significance to 'similarity' and 'dissimilarity' with actuality when evaluating counterfactuals. And we pay more heed to the 'relevant' and 'irrelevant' nature of these similarities. But the notion of 'relevance' is itself developed on purely factual grounds: from the statistical laws governing the actual world.

¹⁰⁶There might be other ways of bringing about the demanded non-contingent relation between the dispositional property and its antidote. To construe such instances in a naturalistic manner is, however, not as straightforward as it might at first appear. But if there are other ways of construing

compound object out of the object possessing the dispositional property and the object which serves as its antidote. We fix the supporting airbag to our china vase and thus compose a compound object consisting of both vase and airbag. And we mix a little haggis into each poisonous pill and thus manufacture a new combi-pill containing both poison and the antidoting haggis. With the antidote thus attached to the object possessing the dispositional property we have created a closed system, as it were, and we can truly say that an antidote is mask-apt once placed in such a closed system: now it possesses the dispositional property to prevent the characteristic manifestation of the original disposition whenever the original disposition is to be triggered. The haggis now is disposed to produce non-poisonous symptoms in anyone ingesting the pill and the airbag is disposed to prevent the china vase from shattering when struck.

However it is still not obvious that the cases so modified qualify as genuine instances of masking. For, with the introduction of compound objects, or closed systems, if you like, it is no longer clear to which entity we should ascribe the original disposition. The semantic difficulty raised by the principle of Variable Strictness is now solved. With the introduction of compound objects it is clear that the subjunctive conditional featuring on the right-hand side of CA is falsified. The carefully airbag-supported china vase would not shatter if struck; and the combi-pill would not kill if ingested. But now it becomes unclear whether the left-hand side of CA remains true. Certainly the china vase, all on its own, is disposed to break if struck. And the poison, all on its own, is certainly disposed to kill if ingested. But on the right-hand side of CA we are not any longer considering subjunctive claims regarding the poison and the china vase all on their own, but rather are considering subjunctive claims regarding the compound objects, i.e. the

masking instances, the following discussion only applies to one particular sort of *masking*, and at the same time the one sort of *masking* that has so far been discussed in the literature, viz., *masking* instances based on composed objects.

china vase with an airbag carefully attached to it; and the combi-pill consisting of both poison and haggis. And it is far from clear that these compound objects possess the dispositional properties in question. At least it is clear from everyday praxis, both linguistically and otherwise, that we ascribe different dispositional properties to simple objects from the ones we ascribe to compound objects comprising these simple objects. We treat a simple object such as an unsupported china vase very differently from the way we treat a compound object comprising both a china vase and a supporting airbag. Is this not because we ascribe different dispositional properties to the two? In fact, it appears that there is a mereological, or, if you like, metaphysical counterpart to the semantic principle of Variable Strictness *viz.*, the mereological principle of Compositionality Strictness:

Compositionality Strictness (Weak)

That object O_1 possesses the dispositional property F does not imply that the compound object O_{1+2} also possesses F —in particular not when O_2 functions as F 's antidote.

The principle of Compositionality Strictness tells us that a true ascription of dispositional properties to an object to a large extent is a function of the mereological status of that object. And, in particular, that a true ascription of a dispositional property well may be a function of whether the property is attributed to an object O all on its own, or to a compound object of which the object O merely is a component. Put another way, according to the principle of Compositionality Strictness, composite objects do not necessarily share dispositional properties with all their constituent parts.

4.5 A Naive Solution

What we could call the *Naive Solution* to the problem raised by masking would thus be to point out that the masker cannot both have her cake and eat it: she

cannot falsify the right-hand side in CA in a way that complies with the principle of Variable Strictness and at the same time satisfy the left-hand side in a way which complies with the principle of Compositionality Strictness.

The proposed naive solution to, or rather dissolution of, problems to do with masking counterexamples to CA might not be the final word on this matter. In particular it might be argued that—principles of Variable Strictness and Compositionality Strictness notwithstanding—we *can* construct masking counterexamples to CA. We can do this, so the argument would go, if we focus attention on a constituent part of a compound object, e.g. if we focus attention on the china vase in the compound object consisting of vase and airbag. The disposition of this object to shatter if struck, it may be argued, does not compromise the principle of Compositionality Strictness since we now explicitly are looking at some property of the vase rather than a property of the compound vase+airbag object. Neither do we compromise the principle of Variable Strictness when we claim the vase would not break if struck since there now is a non-accidental relation between the disposition in question and its masking antidote, the supporting airbag. In order to meet an objection along these lines it might be necessary to strengthen the principle of Compositionality Strictness:

Compositionality Strictness (Strong)

That object O_1 -all-on-its-own possesses the dispositional property F does not imply that O_1 -with- O_2 -attached-to-it also possesses F—in particular not when O_2 functions as F's antidote¹⁰⁷

The principle of Compositionality Strictness so stated implies that an object might change dispositional properties according to its mereological status such

¹⁰⁷ This principle is stronger than Compositionality Strictness (Weak) since it does not explicitly mention two distinct objects (the simple and the compound) but only mentions one object (O_1) in two different modi.

that the vase all on its own might have different dispositional properties from those of the vase with the airbag attached to it. Personally I find such a strengthened version of the principle of Compositionality Strictness plausible. Indeed, it appears to be the only available explanation we can give to the otherwise highly puzzling fact that we treat a vase-all-on-its-own so very differently from a vase-with-an-airbag-attached-to-it. Why should we treat the two differently if we think they possess the same dispositional properties? Indeed, the suggested strengthening might well turn out merely to amount to a linguistic reminder, *viz.*, that when we think and talk about *a vase-with-an-airbag-attached-to-it* we really think and talk about the compound vase+airbag object. With that reminder in place the strengthened and the original version of the principle of Compositionality Strictness become equivalent.

Someone might still wish to press the point that even if this is so, even if 'vase-with-the-airbag-attached-to-it' really refers to the compound vase+airbag object, one is still allowed to consider *a vase-with-an-airbag-attached-to-it* in abstraction from the attached airbag when contemplating which dispositional qualities it possesses and thus conclude that a vase-with-an-airbag-attached-to-it shares dispositional qualities with a vase-without-the-airbag-attached-to-it. And one certainly is entitled to perform this sort of extrapolation. Only should one be careful to exhibit consistency and also consider *a vase-with-an-airbag-attached-to-it* in abstraction from the airbag when reflecting which subjunctive properties it has. The vase-with-the-airbag-attached-to-it considered in abstraction from the attached airbag certainly is disposed to break if struck (left-hand side holds). Likewise we can be pretty certain that it would break if struck (right-hand side holds). For, when considering *a vase-with-an-airbag-attached-to-it* in abstraction from the attached airbag we are considering nothing but the vase-without-the-airbag-attached-to-it. And the vase-without-the-airbag-attached-to-it is surely still a vase-without-the-airbag-attached-to-it when struck in a counterfactual situation

as similar to the actual situation as possible. In such a hypothetical situation it would break.

Another potential line of objection to the naive solution has it that the principle of Compositionality Strictness, although true, is completely irrelevant for construing masking counterexamples to CA, since these counterexamples can easily be constructed without any sort of mereological indexicalisation. Alternatively we simply draw a picture of the compound vase+airbag object. Once that is done we have created a context, as it were, in which it is clear that *the vase*, would not break if struck. This is so since the most similar hypothetical situation in which *the vase* is struck is a hypothetical situation in which the airbag still is attached (just as the contextual representation depicts it as being in the actual world) and thus prevents the vase from shattering. At the same time the contextual representation depicts *the vase* as made of brittle china—everyone can see it is disposed to break if struck. However, we should not get seduced into thinking that we can do magical things with non-linguistic representation which we cannot do with linguistic ones. We cannot.¹⁰⁸ Really, the non-linguistic drawing of a vase with a supporting airbag attached to it depicts a vase-with-an-airbag-attached-to-it just as much as a verbal depiction does. Furthermore, it depicts a vase-with-an-airbag-attached-to-it which surely would not break if struck—but at the same time it depicts a vase-with-an-airbag-attached-to-it which is not disposed to break if struck. True, it is made of brittle china and is thus fragile. But it has also a supporting airbag attached to it and is thus secured against shattering, i.e. disposed

¹⁰⁸Nor should we be seduced into thinking that we can do funny things such as undermining any epistemic tracking counterfactual of the form: were $\neg P$ to be the case, S would not believe P, merely by producing a drawing of a demon on the white board and thus ‘somehow’ changing the context. That drawing, by itself, will at best serve as a reminder that complicated counterfactuals of the form: were $\neg P$ to be the case and were the demon to be around, S would believe $\neg P$, have different truth-values from the simple one. But that is a different matter.

not to shatter if struck¹⁰⁹. Once again we can, if we wish, extrapolate and focus attention on the non-linguistic drawing of the vase-without-a-supporting-airbag-attached-to-it. Here, however nothing is to be found either which was not already present in the verbal depiction of the vase-without-a-supporting-airbag-attached-to-it. What we find is, as expected, an object disposed to break if struck and which would break were it to be struck in a hypothetical situation as similar to the actual situation as possible.

In order to remedy any remaining hesitancy towards the principle of Compositionality Strictness, consider the following bit of ‘mereological calculus’. Surely

(D1) The dispositional property to break if struck; and

(D2) The dispositional property to break if (struck while being supported by an attached airbag)

are two distinct dispositional properties. Some objects possess D1 but lack D2. Likewise it seems obvious that some randomly chosen object, such as a china vase, satisfies the following equivalence, (E):

(E) X possesses D2 iff the compound X+airbag possesses D1

Now, assume the vase possesses D1 but lacks D2. It then follows that the compound vase+airbag object lacks D1 (due to (E)). Hence the vase-without-an-airbag-attached-to-it possesses a dispositional property, *viz.*, D1, which the vase+airbag object lacks. But since the vase+airbag object is nothing but a vase-

¹⁰⁹To be sure, ‘fragile’ and ‘disposed to break if struck’ are two distinct dispositional properties. Different properties whose extensions have a large degree of, although not complete, overlap. This essential distinction will be the main objective for the following section.

with-a-supporting-airbag-attached-to-it, this is an instantiation precisely of the possibility which the principle of Compositionality Strictness announces—the possibility *viz.*, that X-without-its-antidote-attached has dispositional properties distinct from X-with-its-antidote-attached

4.6 Lewis on Naive solutions

The considerations regarding masking just rehearsed have some impact also on altering phenomena. In fact, it seems that the naive solution would be directly applicable to instances of altering. Everything we have said about mask-apt antidotes does, one would think, also apply to altering-apt antidotes (entities that are capable of bringing about genuine altering counterexamples to CA). That is, in order to construct (one sort of) altering counterexample to CA where the right-hand side subjunctive conditional is undermined but some dispositional property still flourishes, our antidote, the object responsible for bringing about the altering, must be *disposed* to *alter* the dispositional property in question whenever the latter is triggered, and not merely happen to do so on particular occasions. And, so the story goes, it can only be so disposed if attached to the original object, i.e. if we operate with compound objects.¹¹⁰ But, as the principle of Compositionality Strictness brings out, objects with their altering antidote attached cannot be

¹¹⁰Martin's 'electro-fink' being a case in point. An electro-fink could for example be a conductor made of thermo-metal (a metal which conducts when it, as now, has its atomic bindings alpha-structured but which immediately swaps to the non-conductive beta-structure if subjected to heat) with a small electrical heater attached to it. Only with the heater thus attached will the heater, non-accidentally, bring about the altering whenever the electro-fink is subjected to current. See the discussion of electro-finks in Chapter III, section 3.3). In other cases of altering, such as the Johnston/Wright chameleon, it is harder to identify and localise the altering entity (this, to be sure, does not mean that there is no identifiable and localisable altering entities operating in such cases).

attributed the same dispositional properties as those objects all on their own. Falsification of the right-hand side thus implies falsification of the left-hand side too. However, Lewis has explicitly denied that such a solution should be available for altering cases. This is rather surprising since he appears to advocate something similar to the naive solution sketched above when dealing with masking phenomena. In the following we will look at this affinity between the naive solution sketched and Lewis' comments on masking.

According to Lewis, masking only raises difficulties for CA when CA is applied to specific dispositional properties such as poisonousness, fragility and inflammability (1997a, pp 152-4). His suggestion is accordingly straightforward: that we should simply specify the dispositional *explanandum* in *general* terms, as for example.

(D3) The dispositional property to cause poison symptoms if ingested

in order to avoid masking phenomena. Why should the formulation of the *analysandum* in general terms, like D3 as the disposition to R if S, rather than in specific terms as simply 'poisonous', make any difference to the success of the CA analysis in the light of masking phenomena? Since Lewis does not consider masking to be any serious threat to CA he does not use much ink expounding his ideas. Here is what he says:

Our chosen *analysandum* [something *x* is disposed to give response R to stimulus S iff] has another advantage: generality. Suppose instead we had taken some particular example of a dispositional concept: the concept of a poison, say, or the concept of fragility or the concept of a lethal virus. A dispositional concept is the concept of being disposed to give such and such response to such and such stimulus. So the first problem we face in analysing any particular dispositional concept, before we can turn to the more general questions that our particular example was meant to illustrate, is the problem of specifying the stimulus and the response correctly.

We might off-hand define a poison as a substance that is disposed to cause death if ingested. But that is rough: the specification both of the response and of the stimulus stand in need of various corrections. To take just one of the latter corrections: we should really say ‘if ingested without its antidote’. [] One who is prepared to speak of masking might stay with the simple definition of a poison as a substance disposed to cause death if ingested, but might say as well that the disposition of poison to kill is masked by antidotes. (1998, pp. 152-3, my square bracket)

The thought here seems to be that the poisonous property of some object, such as our poisonous pill, may be masked by the effect of something akin to our attached haggis pill.

One major initial difficulty which faces us if we opt for a dispositional analysis of some particular property such as poisonous (rather than the general dispositional property D3) is thus to specify precisely *which* property it is we are attempting to analyse. Obviously it is not simply the property D3 since ‘poisonous’ and D3 are divergent in extension. Lewis proceeds by sketching two possible strategies one could adopt in order to solve this initial difficulty. Either one decides that D3 captures the meaning of ‘poisonous’, i.e. one decides that D3 and ‘poisonous’ are equivalent and tries to explain away the apparent divergence in extension with masking which subsequently has to be accounted for. Or one attempts to give a more fine-grained definition of ‘poisonous’ as:

(D4) The dispositional property to cause poisonous symptoms if ingested without its antidote

Both strategies are problematic, to put it mildly, according to Lewis. The former strategy is problematic since it leaves us with the vexatious theoretical burden of

ruling out ‘masking’ phenomena.¹¹¹ Furthermore, even if this challenge can be met, the embarrassing question remains whether we initially opted for D3 as definiens for ‘poisonous’. Would it not have been equally justified to opt for D4, or some other D_n for that matter, and so rule out masking and mimicking instances relative to *that* property? In Lewis’ words:

even if we say that the poison has the disposition spelt out in [D3] and we say as well that this disposition is masked by antidotes, do we not still want to say that the poison has the further disposition spelt out in [D4]? (Lewis 1997a, p. 153)

If no good answer is available here it seems to be a completely random project, and thus rather pointless, to identify specific properties with any of the general candidates (i.e. candidates satisfying the ‘disposed to R to S’ format) and subsequently rule out masking relative to that candidate rather than relative to some other candidate. ‘Poisonous’ can only be equivalent to both D3 and D4 (and D_n) if these various D-candidates are merely potentially divergent in extension. But, and this is precisely the root of the problem: it is demonstrable that they are *actually* divergent in extension.

This points us in the direction of the alternative strategy sketched above, namely to opt for a complex specification of ‘poisonous’ which accounts for all potential masking mechanisms. But this too seems to be a burdensome task. The reason is that many sorts of mask-apt antidotes may prevent some poison from displaying its expected manifestation. In fact it might well be that there are infinitely many such mask-apt antidotes. So the difficulty stated can only be circumvented if it is possible to incorporate a complete list accounting for all potential sorts of mask-apt antidotes in our specification of the dispositional property. However, it is

¹¹¹A highly challenging task indeed, and, furthermore, a challenging task which the best approaches so far have failed to meet—as pointed out in Wright 1991a.

highly doubtful whether such a complete specification can be provided in a substantial manner. As pointed out by Wright, the temptation is to render CA vacuous by inserting an ‘unless *somehow prevented*’ clause:¹¹²

D5 The dispositional property to cause death or illness when ingested unless somehow prevented from doing so

In appealing to the catch-all clause ‘unless somehow prevented’ in our definition of ‘poisonous’, everything (and nothing) becomes poisonous. The Cornflakes you had for breakfast might well have caused your untimely death if you had developed a sudden and strong corn flakes allergy. Under normal conditions you do not develop such an allergy. Only the fact that you do not prevents the Cornflakes from killing you.

Furthermore, suppose it were possible to work out a formulation of the highly complex dispositional property of poisonous which accounted for a totality of potential ways the manifestation could be frustrated; and, suppose further, it were possible to do this in some non-*ad-hoc*-ish manner, for example along the following lines:

D6 The dispositional property of some substance to cause-death-or-illness-in-otherwise-healthy-humans-who-haven’t-recently-eaten-haggis-or-in-some-other-way-been-treated-against-poisoning-and-whose-digestion-systems-function-normally-within-a-reasonable-time-span-after-digestion-of-a-suitable-amount-of-the-substance-in-question

this would still only solve half the difficulty. It would state the necessary conditions for ‘poisonous’ but not the sufficient conditions. An ice-cream might

¹¹²Wright 1998b.

thus, for all that have been said, possess D6. This will be so if it has a uniquely wonderful taste—such a wonderful taste that the profound experience of ingesting the ice-cream releases certain chemicals in the brain which, in turn, trigger a strong depression terminating in a suicide. In fact, there are numerous ways in which something can have D6. And usually it is only one of those ways, one *response-specification* as Lewis puts it¹¹³, that is built into a particular property such as ‘poisonous’. And, again, it is no easy matter when defining ‘poisonous’, to be explicit about which of these response-specifications is being singled out.

Now, Lewis’ suggestion is that we should circumvent the difficulties involved in either of the strategies just discussed by opting for a generally specified *analysandum*, such as D3, in our CA analysis rather than the particular dispositional property, ‘poisonous’.¹¹⁴ Lewis’ advice is, in other words, not to be too concerned if certain non-poisonous objects, such as our wonderful ice-cream, possess D3 while certain poisonous objects do not. Rather, we should focus our attention on the crucial theoretical task of elucidating upon what it actually implies to possess D3.

Lewis’ suggestion is, in effect, that we reverse the order of analysis. According to our analysis, we first pick some particular dispositional property such as ‘poisonous’ as our *analysandum*; and then proceed to identifying poisonous with

¹¹³‘Alternative response specifications’ in Lewis’ terminology being equivalent to our notion of ‘mimicking’. Where we say a vase’s fragility is merely mimicked if striking causes it to break in some unusual manner (bits of dynamite implanted in the glass, say) Lewis says we have built in an derivative response specification in the concept ‘fragility’.

¹¹⁴Cf. Lewis 1997a, p. 152. That Lewis sees his suggestion about general specifications as a solution to the difficulties just sketched is vindicated by the fact that he enters his discussion with a *reductio ad absurdum* strategy: ‘Our chosen *analysandum* [something O is disposed to give response R to stimulus S iff] has another advantage: generality. *Suppose instead we had taken some particular example of a dispositional concept: ...*’

D3. But this leaves us with the difficulty of masking counterexamples which we either, somehow, must rule out subsequently, or, preventively attempt to avoid by specifying ‘poisonous’ as some highly complex D_n . Lewis suggests that we should simply dodge this unpleasant choice by picking the general D3¹¹⁵ as *analysandum* and starting point for our analysis of dispositional properties.

Why is this suggestion germane to the naive solution? Note first that Lewis’ solution to masking logically implies the principle of Compositionality Strictness. What we should focus on, Lewis tells us, is whether some object possesses the general disposition i.e. D3 and not worry too much whether some non-poisonous objects, such as certain ice-creams, possess that property too. The resolution Lewis advertises is to realise, and accept, that ‘poisonous’ is not the same as the dispositional property to cause death or illness if ingested.¹¹⁶ Similarly, ‘Conductive’ is not the dispositional property to conduct if subjected to current; and ‘red’ is not the dispositional property to appear red if looked at. Some objects possess the first but not the second dispositional property in these pairs. The

¹¹⁵Or some other general dispositional property satisfying the ‘disposed to R to S’ format.

¹¹⁶Poisonous is not the dispositional property to cause death or illness if ingested. Is the dispositional property to cause death or illness if ingested the same as the dispositional property to cause death or illness if ingested with an antidote, or more generally, the same as the dispositional property to cause death or illness if ingested and X? The clear and distinct answer is yes—and no. Yes, if ingesting poison counter-entire-factually implies X. No, otherwise. Another way of putting the same point: yes if X is among our default assumptions relative to poison ingesting. No, otherwise. It is true that specific contextual considerations have an enormous influence on our evaluation of counterfactual assertions. But it is equally true that we have a set of default assumption relativised only to the counterfactual antecedent. It is in virtue of these default assumptions that counterfactuals, *simpliciter*, can be evaluated at all. To deny the existence of default assumptions is to deny counterfactuals *simpliciter* their place as meaningful expressions in the language.

significant question is: which objects? And the answer, of course, is: objects whose dispositional properties are ‘masked’: objects with a mask-apt antidote attached to them.¹¹⁷ According to Lewis, objects with their antidote attached to them have thus the former but not the latter dispositional property from our lined-up pairs. Objects *without* any mask-apt antidote attached, on the other hand, possess both dispositional properties from these pairs. So Lewis’ resolution is, in fact, just an alternative way of stating the principle of Compositionality Strictness:

Compositionality Strictness (Strong)

That object O_1 -all-on-its-own possesses the dispositional property F does not imply that O_1 -with- O_2 -attached-to-it also possesses F —in particular not when O_2 functions as F ’s antidote

The naive solution and Lewis’ solution are thus based on the very same theoretical edifice. Lewis is of the opinion that this edifice only bears the burden of masking remedies. However, as we shall see, it also has the strength to bear the burden of altering remedies.

4.7 Mimicked Masking Contexts — Quantifying In

Before implementing our naive solution on altering counterexamples to CA we will consider one last objection. In the foregoing we made a swift move from the intuitively plausible Weak Compositionality Principle to the slightly more controversial Strong Compositionality Principle, and we justified this move with the ‘linguistic reminder’ that, in discussing some object O with its antidote A

¹¹⁷Again, there may be other (so far undetected) ways of construing ‘*masking*’ cases than the one based on mereological surgery which has been the focus of our discussion. If so, there are other objects that have one, but not the other dispositional property from the pairs lined up. That, however, would not be essential for the present argument.

attached to it, what we are really discussing is O', the composite object consisting of A and O—a distinct object possessing dispositional properties distinct from those of O.

Consider the following scenario: Jones is about to jump from an aeroplane. Jones, all on his own, is disposed to die if he does so. However, he is wearing a dispositional antidote to this disposition, a parachute, on his back. So the 'closed system', the parachutist consisting of both Jones and parachute, is not disposed to die by jumping. Our 'linguistic reminder' amounts to the claim that Jones with an attached parachute just is the parachutist. And since the parachutist is not disposed to die by jumping, neither is Jones with an attached parachute. However, by the same token, one could also insist that Jones with an attached parachute is identical with Jones all on his own. And since the latter is disposed to die if he jumps, so is the former. The truth is that both are perfectly appropriate interpretations, and, accordingly the claim:

- (4) Jones with an attached parachute is disposed to die if jumping from the aeroplane

is ambiguous between the two readings. It takes either the parachutist or (mere) Jones as grammatical subject. In the former case it states something false, *viz.*, that the parachutist is disposed to die if jumping. In the latter case it says something true, *viz.*, that Jones who, surely, possesses the typical human dispositional property to succumb if subjected to the consequences of a 5,000 foot free drop, and will retain that dispositional property if situated in a particular context, is in this case placed as constitutive component in the closed Jones-parachute system. The second reading takes the simple, non-compound object as the grammatical subject. Parsing this way will retain varying degrees of plausibility according to the case in question. In the parachute example, it yields a minimal degree of

plausibility, but in other cases, it yields less questionable readings. For example, in our Pill case,

(5) a poisonous pill with an attached antidote is disposed to kill if ingested

is more comfortably read as stating that a poisonous pill is disposed to kill if ingested—even when placed in an antidote context. In any case we cannot, without further ado, deny the masker her right to opt for this interpretation when evaluating the right-hand side of CA in her alleged masking cases. And thus interpreted the masker has a constitutive counterexample which has identical grammatical subjects on both left-hand side and right-hand side in CA. However in thus insisting on her preferred interpretation for the sake of uniformity regarding grammatical *subject*, the masker creates a new equivocation on the grammatical *predicate*. The structure of her counterexample is:

(6) O with A attached is disposed to R if S

In forcing the separation of the ‘with A attached’ phrase from the grammatical subject, in order to retain O as subject, the masker thereby attaches it to the grammatical predicate. On the face of it the masking grammar can either be interpreted as:

(7) **O with A attached** is disposed to R if S, or

(8) **O** with A attached is disposed to R if S

(8) is equivalent to:

(9) O is disposed to R if (S & A)

where the disposition to R if (S & A), to be sure, is a distinct disposition from the disposition to R if S. The masker wishes, of course, to get rid of the ‘with A attached’ phrase altogether, or, at worst, bracket it, and thereby indicate that it serves grammatically to modify neither the subject, nor the predicate. However, neither option is available: it cannot be presupposed that the phrase is tacitly assumed in the CA analysis and thus needs no explicit mention. It is agreed on all hands that the CA analysis is sure-fire when employed to analyse dispositional properties of simple objects, or, as we put it, an object set in a complex system where suppression of the expected manifestation is a statistically rare phenomenon—always ascribable to some specific accidental state of affairs. But this is just to agree that the claim:

(10) O is disposed to R if S iff were S to obtain, R would obtain

simpliciter, is true. In making play with closed systems, in the form of composite objects, the masker is, in effect, prefixing an ‘antidote’ operator to CA:

(11) with antidote A attached (O is disposed to R if S iff were S to obtain, O would R)

The crucial question is how this antidote operator is supposed to distribute over a biconditional. The masker obviously wishes the right-hand side to come out as:

(12) were S to obtain O with an antidote attached would R

which, admittedly, is false. But how about the left-hand side? Merely to stipulate that the left-hand side is unaffected by the antidote operator is question-begging. Which kind of operator would that be that merely distributes into the right-hand

side of a biconditional?¹¹⁸ The suggestion conveyed by the ‘linguistic reminder’ was that the antidote operator works on O on the left-hand side such that the left-hand side reads:

(13) O-in-an-antidote-context is disposed to R if S

which is equivalent to:

(14) O' is disposed to R if S

where O' is the composite, antidoted object. Alternatively it might be that the antidote operator works on S, such that the left-hand side reads:

¹¹⁸Presumably the same suspicious sort of operator as the ‘demon operator’ that the epistemic contextualist sometimes prefixes to the biconditional (S knows P iff S’s evidence rules out all relevant $\neg P$ possibilities). The epistemic contextualist takes his prefixed demon operator to operate only on the right-hand side of such conditionals, especially to widen the scope of the ‘all relevant’ (the idea being that the demon operator add hitherto excluded demon scenarios to the class of relevant scenarios). Hence, according to the contextualist the question whether an epistemic agent knows something on the basis of some particular evidence e depends on the context in which the biconditional is evaluated: if it is evaluated in a demon-free context she knows. If it is evaluated in a demon context she does not know. But, as pointed out in Schiffer 1996, if this is a correct view of the semantics underlying knowledge claims, how could there ever be such a thing as a sceptical puzzle? The puzzle arises because common opinion has it that we know a lot while the sceptic apparently can demonstrate that we know next to nothing. If the contextualist gives a true picture of epistemic semantics, such a tension between common opinion and the sceptic is to be likened to the tension between two interlocutors, disagreeing in the course of a long-distance telephone call as to whether or not it is raining.

(15) O is disposed to R if S-in-an-antidote-context

A third, and final, option is that the operator works on R such that the left-hand side reads:

(16) O is disposed to R-in-an-antidote-context if S

In the Pill case, the three formulations would read: “the combi-pill is disposed to cause death if ingested”, “the poisonous pill is disposed to cause death if ingested with its antidote” and “the poisonous pill is disposed to kill an antidoted person if she ingests it”. All three formulations are false and the original counterexample to CA can thus not be sustained.¹¹⁹

4.8 *Altering Altering*

We can now proceed to a discussion of altering, instanced where the right-hand side and the left-hand side in CA come apart due to an alteration in the object in question *just* after the dispositional property has been triggered, and an alteration furthermore which implies either destruction or creation of the dispositional property in question. In the former case the object will thus possess the relevant disposition, but would lose it were it to be triggered, and, crucially, would lose it *just* before the manifestation would show. In the latter case the object would not possess the relevant disposition but were the characteristic stimulation to take place the object would gain it *just* in time to display the characteristic manifestation. In both cases the right-hand side and the left-hand side in CA come apart—the bringing about of the stimulus causes the object to perform differently from the way CA predicts.

¹¹⁹This point is developed further in Gundersen 2000a.

Here are three examples: consider a circuit breaker.¹²⁰ It consists of not only a conductive circuit but also an attached disconnective mechanism which breaks the circuit whenever it is subjected to an electrical current of sufficiently high voltage. The circuit, in itself, is disposed to conduct when subjected to high voltage currents. The other part, the breaker, on the other hand, is disposed to cut off the circuit whenever it conducts high voltage currents. The manifestation of the circuit-disposition—the conduction of high voltage currents—serves thus as stimulation of the breaker-disposition. The breaker-disposition, in turn, manifests itself by destroying the circuit-disposition by altering one of the circuit's categorical B-properties—connectedness.¹²¹ This circuit breaker is thus *alterably* disposed to conduct high voltage currents. The underlying categorical B-property¹²² in question, connectedness, changes under S to become disconnectedness. Accordingly, the circuit breaker loses its disposition to conduct whenever S obtains.¹²³

Another example of altering dispositionality is provided by the chameleon which, sitting in the dark, is grey-brown but, when brought under observation-apt conditions, including some sort of illumination, will turn into a greenish shade in

¹²⁰Also known in the literature as an 'electro-fink on reverse cycle'. The example is, initially, due to C. B. Martin in his 1994, but refined by Lewis in Lewis 1997a.

¹²¹If you do not like the idea of 'connectedness' being a B-property, just modify the example: let the circuit be made of 'thermo-metal' and let alpha-bonding be the relevant B-property.

¹²²That is, the underlying categorical (as opposed to dispositional) property cf. the discussion of B-properties in Chapter III, section 3.2.

¹²³A concern about timing naturally comes to mind here: does the O-manifestation trigger the M-disposition before or after *complete* O-manifestation? For the sake of argument we will legislate M to be sufficiently sensitive to trigger just before full O-manifestation. In Lewis 1997a, the idea is discussed, and, I think rightly, rejected, that this question of timing can save CA from its apparent *altering* counterexamples.

order to adapt to its present environment (this, of course, it is attempting to accomplish all the time—only it is not quite adept enough to practice this skill in the dark).¹²⁴ Is it disposed to appear grey-brownish if looked at by Camilla? Again we have a compound system. This time consisting of two compounds: the chameleon skin; and the rest of the chameleon. The skin is disposed to retain its colour, grey-brown, under conditions of observation. However, ‘the rest’ is disposed to destroy that very disposition by turning the skin greenish under conditions of observation. We are thus presented with the alterable disposition to appear grey-brownish under observation conditions.¹²⁵

A final illustration of altering is borrowed from the medical world. The human body with all its fine-tuned functionings is fatally vulnerable to bacterial or virulent invasions. In fact the essential biological processes would cease to function within a few hours after an infection—had it not been for the wonderful capacity of the immune system promptly to produce sufficient white corpuscles to fight the invaders. Due to the prompt production of white corpuscles, the human

¹²⁴This example is due to Wright and is inspired by a slightly different chameleon story, namely Johnstons Chameleon case which appeared in Chapter III as an illustration of altering dispositionality.

¹²⁵If having a certain colour is a dispositional property—the disposition to appear that colour to observers under C—then we should say about the grey-brown chameleon in the dark which is disposed to change the colour it has under C, that it is disposed to change disposition under C, that is, change from having the disposition to appear grey-brown to observers under C to having the disposition to appear green to observers under C. It is thus disposed, under C, to gain the disposition to appear green to observers under C. But what relative to one categorisation of the B-properties, is a disposition, under C, to get a disposition under C to appear green to observers, is, relative to some other categorisation of the same B-properties, simply a disposition to appear green to observers under C. A grey-brown chameleon is thus disposed to appear green to observers under C. The problem is, can we then maintain that having a particular colour is a dispositional property?

body loses its nerve-racking disposition to expire when infected shortly after this disposition is stimulated and, in any case, sooner than the disastrous manifestation displays.

Initially, it is worth observing that in these cases of altering, the object O in question must possess some intrinsic property B in virtue of which O can be said to possess the dispositional property—a ‘causal basis’ as it is often termed in the literature. An *altered* disposition of fragility of some object, a vase say, is due to the causally operative B-property of being molecular beta-structured, say, as opposed to alpha-structured which is the B-property underlying the disposition of solidity. Having defined the underlying B-property, we can then go on to say that the vase’s disposition to break is *alterable* because the vase swaps from its beta-structure mode to its alpha-structure mode whenever the disposition is triggered, i.e., when struck. The vase thus fails to satisfy the right-hand side of CA. However, we claim the vase has the dispositional property of fragility, we adhere to the B-properties. The vase is still fragile, we envisage, because it, after all, has the fragile characteristic beta-structure. The intuition is that, had the vase retained the relevant B-property, it would have broken. Hence, the vase is fragile. When there is no such underlying categorical B-property to adhere to, talk about alterable dispositions does not make sense. The point is, really, rather simple: in order for a disposition to be altering, some alteration, and *a fortiori* some alterable B-property, must be to hand.¹²⁶

¹²⁶Another way of putting the same point: construing cases involving altering dispositions is amenable to certain rules. We cannot simply go Berkelyan and suggest, for example, everything is invisible, or non-existent, except when observed by someone, and that everything thus possesses the altering disposition of invisibility. The altering B-properties must be specified in order to give us an understanding of how this alterable disposition is supposed to work.

Nowadays, it is a widely held assumption that *every* disposition can be associated with underlying B-properties.¹²⁷ However, this is still a controversial thesis. And even if true, alterable dispositions call for more than the mere existence of categorical B-properties: It must be possible to offer some account of how to *identify* the relevant B-properties. To illustrate, consider an instance of the alterable disposition of being courageous, i.e. to act courageously when encountering a dangerous situation. Somehow it does not sound very credible to claim there are cases involving genuine courage—but genuine courage that is frustrated whenever a situation calls for courageous behaviour. That seems more like a paradigm case of cowardice. The reason that we have difficulty in conceiving of this case as one of altering is that no obvious candidate springs to mind for an underlying B-property that is destroyed when the disposition is triggered. And, to be sure, the difficulty is not that we cannot conceive at all of any relevant B-property underlying and, to some extent, causally determining courageous behaviour. There may well be such courageous B-properties. There may be particular genotypes characterising courage. Or there may be a higher than average concentration of certain hormones flowing around in courageous people. Or just some neuro-biological defect in the corner of their brains responsible for rational agency—or what have you. The point is rather that it is hard to conceive of any of these B-candidates as being the slightest bit susceptible to undergo changes when the disposition is triggered. For example, why should we think that the genetic key for courage (if there be any such) suddenly undergoes radical changes whenever a bull, say, is approaching at disturbingly high speed?

In cases of response-dispositions,¹²⁸ such as colour competence, where manifestation essentially involves some mental activity,¹²⁹ it is doubtful whether

¹²⁷By *inter alia* Armstrong 1973, Prior *et al.* 1982, Jackson 1996 and Lewis 1997a and 1997b.

¹²⁸As defined in Chapter II.

such underlying categorical B-properties exist—and even more doubtful whether, if they do, they can be said to be associated with triggering of the disposition in such a way that an altering scenario is rendered plausible. In case of an *external* response-disposition, for example, of brownness, the property of having surface corrugation characteristic for brownness naturally comes to mind as a B-property candidate for altering. It is in virtue of this property that for instance the above-mentioned chameleon example bears its plausibility. We can readily conceive of ‘the rest’ of the chameleon as somehow manipulating that property whenever the beast is having its dispositional skill tested. But what is the corresponding underlying B-property for the *internal* counterpart disposition—brown-competence (one’s disposition to see brown things as brown whenever one comes across them)? We can point to a causal chain starting with a certain sort of electromagnetic radiation of such and such a corrugated surface and, via a particular sort of deflection and some neuro-physiological processes, issuing in a certain brain state. Whenever this causal chain is activated and successfully completed, the agent is in a certain brain state.¹³⁰ But which categorical B-property is responsible for the last transition from this brain-state and to the sensation of brownness? Any candidate that comes to mind, including possession of a ‘brown-gene’, seems merely to influence the complex causal chain between stimulation of the disposition, looking at the object, and the brain state. And the difficulty involved in identifying such an underlying B-property is reflected in the fact that it does not make much sense to propose the existence of an *alterable* disposition of brown-competence: a genuine disposition to see brown things as

¹²⁹As opposed to external response-dispositions where the response-task can be performed equally well—and sometimes better—by apparatus.

¹³⁰By ‘certain brain state’ we only understand merely the theoretically innocent ‘a brain-state which, as a mater of fact’ issues in such and such a sensation. Nothing as presumptuous as a ‘brown brain-state *type*’ is assumed.

brown—except when one looks at them, in which case one sees them as green!¹³¹ The significance of this point will become apparent in next chapter where we will discuss the case of altering specifically in the context of epistemic dispositional.

Now, as mentioned, the ‘naive solution’ developed in this chapter as a response to the challenges masking raises for CA also bears some plausibility as a potential remedy for difficulties raised by cases of altering. Like masking, genuine cases of altering must, in order to meet the constraints defined by the principle of Variable Strictness, secure a non-accidental connection between the disposition and the associated altering mechanism. In the above cases, the altering mechanisms, or dispositional ‘immune systems’, as we might phrase it, were respectively the ‘breaker’ from the circuit breaker, the ‘rest’ of the chameleon, and the biological immune system. The behaviour of the disposed objects—the circuit, the chameleon-skin and the vulnerable body—detached from their dispositional ‘immune systems’ comply nicely with CA. A brown chameleon-skin, all on its own, would appear brown if observed;¹³² the circuit in the circuit-breaker, all on

¹³¹Which is a great pity! If it were feasible we could have provided for some beautiful philosophical examples of internal and external pairs of response-dispositions with synchronic altering. Green things turning red under C — in tandem with green-sensibility being transformed to red-sensibility. These kinds of example become, though, feasible when it comes to masking and mimicking as we saw in Chapter III. In Chapter V we shall return to this possibility of synchronised masking.

¹³²One rather disturbing consequence of this particular claim is that the skin as such, but not the whole creature, is disposed to appear brownish when the disposition is triggered. This is a puzzle since the skin and the chameleon also, by and large, are disposed to appear same-coloured. However, one can easily respond to this puzzle by pointing out that in mentioning the skin above, what we really mean is the chameleon-skin, i.e. the skin considered in tight relation to the remaining chameleon and its complex of biological functionings. Our chameleon example is, however, inviting us to consider the skin as such as an independent entity. And the skin all on its

its own, would conduct if subjected to high voltage currents; and a human body with a malfunctioning, or absent, immune system would expire if infected. Furthermore, were some artificial immune system to immunise the dispositions in question on some random basis and thus produce a statistically divergent, isolated instance of a genuine disposition that, on this particular occasion became immunised sooner than the expected manifestation displayed, we would still like to say that the object in question, generally, would produce the expected manifestation if stimulated. For example, were some unfortunate person with a malfunctioning immune system to ingest a piece of Danish Blue cheese with an artificial vaccinating effect just at the fateful hour she suffered her otherwise fatal infection, we would still like to say she would die were she to be infected. That she was blessed with extraordinarily good fortune on this one occasion does nothing by way of dispensing with her sad karma: were she to be infected she would die.

In order to generate a genuine case of altering one must thus create a closed system consisting of the disposed object and its immunising entity. The obvious manner to do this is by creating a complex object but with the appearance of such a complex object it becomes unclear *which* entity it is that fails the CA test. The composite object would not display the manifestation in question, but nor does it possess the relevant disposition. As illustrated by the principle of Compositionality Strictness, the compound object does not share dispositional properties with all its subparts. The circuit breaker is not disposed to conduct high voltage current although one of its components, the circuit, is. In fact, it has been

own is certainly differently disposed from the chameleon-skin, i.e. the skin considered in tight relation to the remaining chameleon and its complex of biological functionings.

Likewise the skin, as such, is disposed to remain motionless when a dangerous situation accrues. Nevertheless, when we conceive of our chameleon as liable to flee if frightened, we conceive of the chameleon-skin as disposed to accompany the rest of the beast in its retreat.

designed with the particular disposition to *prevent* conduction of high voltage currency. Nor is the human body disposed to expire if infected just because the subpart consisting of all the bodily organs, save the immune system, is so disposed. Likewise the chameleon does not by any means share the disposition of its skin permanently to appear same-coloured. On the contrary, the entire creature is disposed to exhibit its camouflaging skills successfully and thus to swap colour when observation conditions (daylight) obtain.

At this point we could, if we wished, run through the same potential objections as those raised earlier in connection with masking context and quantifying in. However, that would be a mere repetition, as the dialectic on these matters follows a similar pattern to that which emerged in the discussion of masking. Instead we will attend to Lewis' response and discuss to how large an extent it agrees with the Naive solution in its treatment of masking and altering respectively. Whereas Lewis, as mentioned above appears to be in accordance with the Naive solution as the appropriate remedy for masking, he explicitly denies that it may have any bearing on the difficulties raised by altering. Here is his response to the fictive interlocutor, the *resister*, who comes up with just that suggestion:

A different line of resistance suggests that if something is [alterably] disposed to give response [R] to stimulus [S], what it really has is a compound disposition. It has a state that at least resembles a disposition to give response [R] to [S]. Our resister, since he accepts [CA] will think it inaccurate to call this state a disposition. (I shall signal his terminological scruples with inverted comas.) At any rate, this first 'disposition' is embedded in a second disposition. The thing is disposed to lose the first 'disposition' in response to [S].

Now the resister is struck by the difference between the first 'disposition' all by itself and the first 'disposition' when it is embedded in the second. He implores us not to be over-impressed by such similarities as there are, and instead to heed the difference the second disposition makes to the overall dispositional character of the thing. When we say that the thing is disposed to give [R] in response to [S], he thinks we are misled by thinking of the first

‘disposition’ in abstraction from the second. (1997a, p. 147; my square brackets)

However, a resister can conduct his response much better than that. A resister need not press the point at all whether the embedded disposition really is a disposition or merely a ‘disposition’. Nothing as desperate as that is called for. A resister can quite happily grant that the embedded disposition really is a disposition. With findings to hand accruing from reflections on the principle of Variable Strictness and principle of Compositionality Strictness, he need only stress that one should be careful to specify the object we subject to the CA analysis. Once the bearer of the disposition is properly identified, and consistently referred to on both left-hand side and right-hand side of CA, the apparent altering counterexamples vanish.

Lewis himself grants the resister’s line of thought some plausibility but leaves the matter rather undecided. To the question whether or not the ‘resister’ ultimately can provide the theoretical means for solving the challenges raised by altering, Lewis replies:

That may be so, or it may not, in the sort of case the resister has in mind. (I myself think it is not so).¹³³

Lewis can afford this laxity since he believes the *alterer* has the resources to defeat CA even if the resister is granted his points regarding composite dispositionality. To this purpose he offers the following case of altering which, admittedly, is untroubled by any considerations rehearsed so far:

A sorcerer takes a liking to a fragile glass, one that is a perfect intrinsic duplicate of all the other fragile glasses off the same production line. He does nothing at all to change the dispositional character of his glass. He only

¹³³*Ibid.* Unfortunately Lewis does not comment upon why he himself does not think this is so.

watches and waits, resolved that if ever his glass is struck, then, quick as a flash, he will cast a spell that changes the glass, renders it no longer fragile, and thereby aborts the process of breaking. So his [altering] fragile glass would not break if struck—but no thanks to any protective disposition of the glass itself. Thanks, instead, to a disposition of the sorcerer. (*Ibid.*; my square brackets)

And he concludes his discussion of the matter by noting that “To refute an analysis, one counterexample is all we need.” That may be so. However, to pick just this particular candidate for accomplishing that refutation strikes one as a little optimistic. In fact, a suitable reaction to this case, and generally to cases involving supernatural agency, is to follow Blackburn’s advice and respond to it simply by inserting a ‘naturalising’ clause in CA (1993c, p. 266):

(CA') Object O has the dispositional property to R if $S \leftrightarrow O$ is *naturally* such that it would produce R were S to obtain

the effect being to restrict the explanatory power of CA to natural phenomena, i.e. to phenomena with which science—rather than magic—engages itself. Nothing much is lost by this concession anyhow since it surely must be the fundamental assumption of any intellectual enterprise aiming at some degree of generality. Only magic would be interested in accounting for such remarkable, or should we say miraculous, incidents as angelically protected fragile objects, transcending any scientific comprehension of a law-governed universe. In fact, it is doubtful whether such a naturalising clause is called for in the first place. Is object O not exactly as O naturally is? If so, the inserted ‘naturally’ seems superfluous. The most obvious way with the fragile object in the Sorcerer case is simply to insist that it would, as a matter of fact, break if struck. The Sorcerer, since non-existent,

would let it down.¹³⁴ In any case it is rather surprising that Lewis lends so much credence to sorcerers in his dialectical strategy against the resister.¹³⁵

As mentioned, there may in addition to those responded to here, be other plausible ways to construct a closed system in which the non-accidental relation between the disposed object and its immuniser is secured. However, it remains to be specified how these systems are supposed to work and, in particular, why they are indifferent to the defence for CA developed in this chapter. There may or there may not be such cases. One thing is certain, though: Lewis' Sorcerer case is not one of them.

4.9 Summary

In our discussion of masking we have considered two different scenarios: those where the masking takes place in a complex system, i.e. a system characterised by lack of any obvious causal, or otherwise non-accidental connection between the disposition and the operation of its dispositional antidote; and those scenarios where the masking takes place in a closed system, where things are set up in such a way—most notably by making composite objects—that the antidoting mechanism is bound to kick in whenever the disposition is triggered. In complex systems we have argued that the right-hand side subjunctive conditional in CA is not quite so easily falsified as is often assumed in the recent literature on the matter. The Principle of Variable Strictness informs us that it does not suffice to falsify a version of the subjunctive conditional (right-hand side of CA) with a strengthened antecedent; in general falsification of $(P \ \& \ R) \ \Box \rightarrow Q$ only implies falsification of $P \ \Box \rightarrow Q$ if $P \ \Box \rightarrow R$ (P and R , in this particular case, being the

¹³⁴Since I wrote this I too have come to believe in the existence of sorcerers. However, I do not think that fact undermines the principal point.

¹³⁵Thinking about it, it is more than surprising: it is outright suspicious.

triggering of the disposition and the presence of some dispositional antidote, respectively). And complex systems just are characterised by an accidental connection between the triggering of the disposition and the frustration of the expected manifestation. According to standard subjunctive semantics, an accidental P Q relation can confer truth on the associated P Q conditional, *viz.*, in cases where P and Q, as a matter of fact, both hold true. We have argued against that view and, further, taken a few preliminary steps towards an alternative subjunctive semantics which is based on a statistical analysis according to which such counterintuitive results are precluded. For those in favour of the standard semantics, we offered the Accidental Clause which will rule out instances of masking in complex systems even under the assumption of standard semantics for subjunctive conditionals.

In closed systems it is obvious that the right-hand side subjunctive clause in CA is falsified. But, still, it is questionable whether the left-hand side is not falsified in tandem. We observed that, even in a closed system, no difficulties are raised for CA when it takes a simple object as subject. Nor is its theoretical adequacy compromised when we pick the entire closed system, the compound object, as grammatical subject in our CA analysis. It is only when we are concerned with the simple object *considered as constituent in a closed system* that trouble beckons. We investigated how it is brought about that we sometimes come to consider an object in that particular way while applying CA. The suggestion was that it is brought about by what we semantically can picture as prefixing an ‘antidote operator’ to CA. We argued that the antidote operator should distribute not only into the right-hand side of the CA biconditionals, as the masker sometimes appears to assume, but also distributes into the left-hand side dispositional claim. We surveyed the three possible ways in which it may be so distributed into the left-hand side and concluded, in all three cases, that the antidote operator falsifies the left-hand side of CA in tandem with its falsification of the right-hand side.

The overall conclusion is that masking counterexamples to CA based on closed-system scenarios as well as complex-system scenarios collapse under the weight of detailed examination. *Mutatis mutandis* the same conclusion applies to instances of mimicking since these are perfectly symmetrical to instances of masking, only with ‘dispositional pretenders’ replacing the dispositional antidotes. In fact, masking with respect to some property F is just mimicking with respect to the complementary $\neg F$ property. Masking a fragile object’s dispositional property to shatter if struck, for example, is just to mimic its solidity—to mimic its dispositional property to hold together if struck.

Furthermore we have seen that the defence developed for CA against masking also provides an efficient defence against the challenges raised by altering phenomena. This is so despite Lewis’ explicit claim to the contrary. Lewis’ discussion of masking appears to be at best in accord with the thoughts underlying the naive solution, but surprisingly Lewis explicitly denies the naive solution as a remedy against altering. His rejection was demonstrated to be insufficiently grounded and, ultimately to depend on a supernatural example. It can, of course, be questioned whether such an example *must* rely on supernatural powers. Is it possible to construct an altering counterexample to CA which neither rely on supernatural powers, nor on composed objects with overall dispositional properties distinct from those of the original object? We have not come across any in the literature so far and have not been able to design any ourselves. We therefore conjecture that no such counterexamples can be constructed.

An issue remains about the fine-grained characterisation of complex and closed systems and, in particular how to draw an exact demarcation between them. For instance, if the antidote is lying on its antitotee without, strictly, being attached to it, are we presented with a closed or complex system? Or if the connection is non-causal but, say, psychological—for example, the kind of psychological tie between someone who functions as antidote for some other person’s disposition to feel depressed: are we presented, then, with a closed system? We can also conceive of

a scenario in which the antidote in question operates on an accidental basis, but on the basis of some sort of predictable accidentalness. As in this case: an otherwise perfectly well-functioning spark plug is sensitive to some specific weather conditions. But science has revealed that just those weather conditions obtain on an overall 4.5 per cent rate. Again the question arises whether we are presented with a complex or closed system. Would the scenario count as an instance where the spark plug's disposition to produce a spark is antidoted now and again on a completely accidental basis; or would it count as a never-failing disposition of an overall closed system to produce a spark on a 95.5 per cent rate? There may be no clear-cut answers to such questions. It may so turn out that the borderline between complex and closed systems is vague and that truth-values of both right-hand side and left-hand side in consequence are indeterminate in these cases. However, such a potential 'grey' area does nothing by way of compromising the theoretical value of CA. Even if we, with Lewis, harshly take one counterexample to suffice for a refutation of the entire theory, it still remains the wisest counsel to allow a theory a few instances of underdetermination.

Another, and even more challenging issue concerns the relation between specified properties amenable to the 'disposed to R if S' format, on the one hand, and simple unspecified properties such as 'poisonousness' *simpliciter* on the other. We have assumed that subjunctive conditionals generally possess objective truth-values and that they do so partly due to the existence of default assumptions about scenarios in which the subjunctive antecedent is realised. This idea provides a clear characterisation of the relation between two specified dispositions: generally the disposition to R if S is identical to the disposition to R if S & A iff A is part of our default assumptions about scenarios in which S is realised¹³⁶. But it

¹³⁶ An example might be helpful here. Take the disposition to break if struck. It is among our default assumption that scenarios in which the striking takes place are scenarios governed by laws

remains an open question how the relation between an unspecified and a specified disposition should be characterised. In fact, it appears that much of the criticism that has been launched against the CA analysis really addresses the intrinsic difficulties involved in translating unspecified dispositions into specified ones. This difficulty may be decisive for certain recent reductivist projects that essentially make play with dispositionality.¹³⁷ However, for our purposes of developing a dispositional theory for warrant and knowledge, the dispositions in question will be exclusively of the specified kind *viz.*, a specified disposition to form a certain belief when rehearsing a certain truth-searching procedure. We will thus leave it to someone else to struggle with this issue of translating unspecified dispositional properties into the ‘disposed to R if S’ format.

of nature. It is not among our default assumptions about these scenarios that an airbag suddenly is activated and thus comes to support the struck object.

¹³⁷For instance for projects aiming for a response-dispositional test applicable on specific core concepts from various discourses such as ‘red’, ‘good’, ‘funny’ or ‘necessary’ and thereby provide for a more sophisticated view on what sort of objectivity governs these discourses. See Johnston 1991, 1992a & 1992b and Wright 1991a, 1992a, 1993b & 1998a.

Chapter V

A Dispositional Theory of Knowledge and Warranted Belief

5.1 Review of the 'Problem Cases'

5.2 A Dispositional Theory of Knowledge and Warranted Beliefs

5.3 Tracker Conditionals versus Tracking Conditionals — A Dispositional Diagnosis

5.4 Certainly Disposed

5.5 A Warrant for Tracking

5.6 Scandalous Scepticism

5.7 Transmission of Warrant

5.8 Aetiological Foundationalism

5.9 Summary

Synopsis

In the light of the findings from Chapter IV, what is the standing of our suggestion that knowledge and warranted beliefs are best comprehended in terms of epistemic dispositionality? In particular, what is the standing of the suggestion put forward in Chapter II and III that knowledge is best considered in dispositional terms—more precisely: as the manifestation of an epistemic disposition to form true beliefs under M-conditions? To recap, the suggestion was that (a) satisfaction of the Tracking Conditionals

(Niii) $\neg P_M \square \rightarrow \neg SB(P)$

(Niv) $P_M \Box \rightarrow SB(P)$

is symptomatic rather than constitutive of an agent's knowledge of P; (b) that knowledge, rather, is constituted by the exercise of a dispositional capability to form true beliefs; and (c) that the attempt to capture this dispositional property in terms of subjunctive conditionals, although providing a good approximation in a wide variety of cases, still comes apart from the real thing whenever masking, mimicking and altering mechanisms are operating. We also saw (in Chapter III) that the intuitive counterexamples to the Dretske–Nozick tracking theory (henceforth the DN Theory) divide smoothly into four categories corresponding to the taxonomy of masking, mimicking and (two kinds of) altering counterexamples to the CA analysis for dispositional properties generally. Against this background it was natural to propose—what I shall henceforth refer to as the *natural suggestion*—that a new, and improved theory of knowledge and warranted beliefs was in the offing either by stating the theory in dispositional rather than subjunctive terms or, alternatively, by improving on the subjunctive formulation in order to account satisfactorily for masking, mimicking and altering instances of epistemic dispositionality.

However, in the course of our search for such an improved formulation in Chapter IV we discovered a strong case that masking, mimicking and altering counterexamples to the CA analysis are merely apparent, i.e. that they dissolve once we pay sufficiently careful attention to certain semantic and metaphysical principles—notably the principle of Variable Strictness and the principle of Compositionality Strictness. This is good news for the CA theorist. And it is now time to examine to what extent it also provides the tracking theorist the means she needs to account satisfactorily for the masking, mimicking and altering of epistemic dispositionality and thus for the four categories of counterexamples to the DN theory. On the face of it there are only two possibilities: either the masking, mimicking and altering counterexamples to the DN theory evaporate in

tandem once we pay sufficient attention to the principle of Variable Strictness and the principle of Compositionality Strictness. Or, these counterexamples to the DN theory survive this scrutiny, in which case the natural suggestion must be flawed: if the Tracking Conditionals provide a good approximation to epistemic dispositionality—an approximation that only comes apart from the real thing in case of masking, mimicking and altering of epistemic dispositionality then these counterexamples to the DN theory must disappear once masking, mimicking and altering has been accounted for. If the problematic cases for the DN theory sustain scrutiny from the principle of Variable Strictness and the principle of Compositionality Strictness it will thus no longer be reasonable to ascribe the problematic cases to the inaccurate subjunctive formulation of an, essential, dispositional concept.

A lot is at stake here and we shall therefore open this chapter by reviewing the problematic cases in the light of the findings from Chapter IV (5.1). As we shall see the findings from Chapter IV do provide some progress for the DN theory. Some of the counterexamples do as a matter of fact dissolve when considered in the light of the findings from Chapter IV. Some, but not all. This *prima facie* disturbing result—the fact that some counterexamples persist—drives us to investigate, in 5.2, whether there might be some demarcating feature distinguishing those counterexamples that dissolve from those that do not. The outcome of this investigation is that there is such a demarcating feature, viz. that the masking, mimicking and altering mechanisms¹³⁸ involved in those

¹³⁸In the light of the findings from Chapter IV it becomes increasingly problematic to apply the notions ‘mimicking, masking and altering mechanism’. Mimicking, masking and altering was defined as instances which undermine the CA interpretation of dispositionality. We have now concluded that there are no such counterexamples to the CA analysis. On the face of it we have thus debarred ourselves from using these notions (at least in the sense just exemplified where it is tacitly supposed that the terms have a non-empty extension). However, for pragmatic purposes we

counterexamples that dissolve are all closely associated with the epistemic agent whereas those masking, mimicking and altering mechanisms involved in the counterexamples that do not dissolve are all closely associated with the worldly state of affairs which the epistemic project is aimed at. In the dissolved problem cases it is the agent and the intervening mechanism that play together and make up a closed system whereas it is the worldly state of affairs and the intervening mechanism that play together and make up a closed system in the unresolved cases.

This finding turns out to be significant. It reminds us of the observation made in Chapter III namely that response dispositions have a dual character: an internal disposition of some agent to respond with a certain belief when confronted with certain entities is always accompanied by an external disposition of those entities to evoke a certain response from the agent. And, as the killer-yellow story brought out, the two dispositions are not always synchronised. A careful examination of this dual character of response dispositionality will reveal that, contrary to what has been assumed so far, the Tracking Conditionals do not represent the epistemic disposition of the agent to get things right about P-matters. Rather the Tracking Conditionals represent the worldly counterpart disposition: the disposition to appear in a certain manner for an epistemic agent.

In the unresolved problem cases the two dispositions—the internal response disposition of the subject and its external counterpart disposition—are not synchronised. Either the agent is not disposed to get things right regarding P (in which case intuitions have it that she does not know that P) but the world is disposed to make her form true beliefs nevertheless. Or, alternatively, the agent is disposed to form true beliefs about P (and intuitions have it that she does know)

will maintain these notions in our vocabulary only bearing in mind that mimicking, masking and altering mechanisms do not generate mimicking, masking and altering counterexamples to the CA analysis.

but the world is unhelpfully disposed to make her form fallacious beliefs about P. And it is precisely in this fact that we find the key to an exhaustive diagnosis of the remaining problem cases. The two dispositions are characterised by two distinct set of subjunctive conditionals. When we first presented the proposal that knowledge ought to be considered as a manifestation of an epistemic disposition to form true beliefs (in Chapter III) we associated the disposition in question with the Tracking Conditionals. The findings in the present chapter uncover that we proceeded a little too hastily. When we associated the relevant epistemic disposition with the Tracking Conditionals we argued that this move was warranted by the CA analysis. But a more careful reading of CA recommends that the relevant epistemic disposition, the *internal* response disposition of the agent, is associated with the *reversed* set of conditionals, the Tracker Conditionals:

Tracker Conditionals

(DISP iii) $SB(\neg P)_{M\Box} \rightarrow \neg P$

(DISP iv) $SB(P)_{M\Box} \rightarrow P$

Once mimicking, masking and altering have been accounted satisfactorily for (Chapter IV), the Tracking Conditionals do as a matter of fact provide an accurate description of an epistemic disposition—only the wrong disposition: the *external* disposition of the world to produce a veracious belief in the agent.

This explains the demarcating feature of the remaining counterexamples to the DN Theory: instances where the mimicking, masking or altering mechanisms are located on the worldly side and thus make up a closed system with the worldly states of affairs. In these cases the overall disposition of the world to produce a certain belief in the agent differs from the corresponding internal disposition of the agent to produce a certain belief about the world and we get the two sort of

counterexamples to the DN Theory that we saw displayed in Chapter III.¹³⁹ And, as we shall see, once this deficiency has been accounted for, the remaining counterexamples to the DN Theory evaporate.

Recognition of the dual character of response dispositionalism and specification of the two corresponding set of subjunctive conditionals does, when combined with the findings from Chapter IV, provide the means for a philosophical diagnosis of the problem cases for the DN Theory and does thus also provide the theoretical basis for an improved formulation of a dispositional theory of knowledge and warranted beliefs. But an additional virtue is that it throws new light on *why* the problem cases for the DN Theory are problematic and, in particular, it throws new light on what lies at the heart of numerous discussions of the classical problem cases (as exposed in Chapter II). In 5.3 and 5.4 we will take full advantage of this theoretical fertility and offer a thorough diagnosis of some of these classical problem cases. We will then proceed to a discussion of the relation between the Tracking Theory (5.5) and our improved dispositional theory vis-à-vis some core topics within epistemology such as scepticism (5.6) and issues to do with Transmission (5.7). These discussions will conclude (5.8) in a rough outline of an epistemic position, Aetiological Foundationalism, which incorporates the resulting views.

5.1 Review of the 'Problem Cases'

But first a review of the problem cases for the DN Theory. Consider first our mimicking counterexample, instances where the Tracking conditionals are satisfied but where the epistemic agent is not warranted. The metaphorical

¹³⁹Unless, of course, a corresponding mimicking, masking or altering mechanism is located in the agent and thus generating a closed system in the agent so the internal and external disposition again is synchronised—more on this possibility later.

characterisation of these instances was given in terms of an arrow dispatched by an unskilled archer who invariably hits his target due to some mimicking device (e.g. a computerised steering mechanism akin to those built into modern missiles) attached to the arrow, or, alternatively some mimicking device (e.g. a strong ‘arrow-magnet’) built into the target. This archer would hit the target were he to dispatch an arrow. But not in virtue of skill. In virtue, rather, of conducive mimicking powers in his close environment. This scenario was intended to represent a situation in which some epistemic agent who lacks the dispositional capability to form true beliefs under M-conditions (conditions where the epistemic method M is being executed) nevertheless invariably hits on the truth when directing beliefs towards the world. This agent would believe P were P to be the case under M-conditions. Also she would refrain from believing P were $\neg P$ to be the case. But this would be due not to her cognitive skill but instead to some mimicking power to bring about P whenever M-conditions obtain or, alternatively, to bring about M-conditions whenever P obtains. The former—M-conditions bringing about P—was illustrated by Flip who investigates whether the light is on next door by the method: tossing a coin.¹⁴⁰ Were the light to be on next door, Flip would come to believe so by executing her coin method.¹⁴¹ This would not be due to Flip’s cognitive capability to get things right, but entirely in virtue of the hidden remote-control mechanism within the coin. The latter—P bringing about the M-conditions—was illustrated by Holo who by observing a hologram of a vase came to believe there was a vase in the box.¹⁴² Were there to be a vase in the box, Holo would come to believe just that by executing this method.¹⁴³ But, again, the epistemic task is not accomplished by Holo’s cognitive capabilities, but rather

¹⁴⁰Cf. The Light case in Chapter III, section 3.4.

¹⁴¹And likewise for the other tracking condition, (Niii) $\neg P \text{ }_M \square \rightarrow \neg \text{SB}(P)$.

¹⁴²Cf. The Holo case in Chapter III, section 3.4.

¹⁴³And likewise for the other tracking condition, (Niii) $\neg P \text{ }_M \square \rightarrow \neg \text{SB}(P)$.

behind her back by the mimicking mechanism—the causal connection between the pressed lever underneath the vase and the activating of the laser projector displaying the vase-hologram.

Now, how do these two cases fare in the light of our two principles of Variable Strictness and Compositionality Strictness? Consider Flip first. Assume the principle of Variable Strictness really is satisfied here, i.e. the mimicking remote-control inside the coin and Flip's execution of her method really is set up in such a way as to constitute a closed system establishing a non-accidental relation between Flip coming to believe the light is on by her coin-tossing method and the light being on. In that case, is not Flip's coin simply to be likened to a sophisticated light switch? But examining a light switch is not, as first assumed, a stupid method to execute when investigating whether or not the light is on. In fact this is a method applied on numerous occasions by competent epistemic agents to whom we readily ascribe knowledge. Your computer is completely dead. What is your first reaction? You look at the switch in order to make sure it is turned on. Or you are about to brew a cup of coffee. First, however, you instinctively make sure the coffee-machine switch is turned on. Or you wish to ensure the light is on before you enter the bathroom. Again, you instinctively examine the switch outside the door before you enter to perform your ablutions. In all these cases your choice of epistemic method is perfectly reasonable and by executing such methods you certainly bring yourself into a situation in which you are disposed to form a true belief: a situation in which your respective true beliefs qualify as knowledge. Why should matters be any different with Flip? Is it because she is completely in the dark regarding the various causal links between the tossing of the coin and the lighting next door? On that score presumably most of us would fail to know—even when observing standard switches. Most of us have not the faintest idea how these widgets function. We have just got accustomed, by ordinary inductive methodology, to the idea that the things are switched on when the widget is in a certain position and switched off when the widget is in a certain other position.

Under the assumption of a non-accidental connection between the coin landing head up and the light being switched on next door we ought to allow Flip the same ignorance concerning the technical details involved in the execution of her method and thus concede she too possesses knowledge as to whether the light is switched on. If, on the other hand, the principle of Variable Strictness is not satisfied, that is, the mimicking remote-control inside Flip's coin does not secure a non-accidental relation between Flip's P-belief and P, then no genuine counterexample to the DN theory has been established. For, if the relation between Flip's belief that P and P is merely accidental—is due to some specific circumstances surrounding this particular case—then Flip would not, generally, believe P were P to be the case. The DN theory thus gets things right in either case.

How about Holo: would we likewise ascribe knowledge to her about the vase's whereabouts once a non-accidental, mimicking tie has been recognised between states of affairs and their holographic representations? Most of us would presumably hesitate to ascribe knowledge to Holo. Despite the non-accidentalness, a suspicion remains that Holo's epistemic relation to the relevant subject matter is still seriously flawed. The exact nature of this suspicion is very precisely expressed by the objection most of us intuitively would come up with in response to this case: "But were Holo to form beliefs on the basis of that method, she would, in a wide variety of cases be led astray!" In particular she would be led astray in a variety of cases indiscernible to Holo from the present one. Watching holograms is like chatting to a compulsive liar. Even wired up in such a way that such a liar always tells the truth in one particular setting, before 9 a.m., say, it still would not count as knowledge if one gathered information from her on one of these favourable occasions. Were one to base one's beliefs on her testimony, one would be positioned very badly epistemically. Note that the same objection would be misplaced in Flip's case. Flip would as a matter of fact form true beliefs were she to apply her method.

Consider next the masking counterexamples to the Tracking Theory. The metaphorical characterisation of these instances was given in terms of an arrow dispatched by a skilled archer who, nevertheless, invariably misses his target due to some masking device and, again, the masking device can either be situated at the dispatched arrow or at the target. Such an archer would, despite his skill, never hit his target. This scenario was intended to represent a situation in which some epistemic agent who possesses the dispositional skill to form true beliefs nevertheless fails to hit on the truth when directing beliefs towards the world. Such a person would not believe P were P to be the case under M -conditions and would believe P were $\text{not-}P$ to be the case. But this would not be due to lack of cognitive skill but instead to some masking mechanism which, triggered by the M -conditions, swaps the truth-value of P , or, alternatively, to some masking mechanism which swaps the M -conditions to become non-conducive for P -investigation. In both cases the masking mechanism blocks the expected manifestation of the relevant epistemic disposition to form true beliefs. The former— M -conditions swapping the truth-value of P —was illustrated by the Reversed Light case where Flip applies some responsible epistemic method to decide whether the light is on next door but where Flip's friend systematically switches the light on and off and thus desynchronises P and Flip's P -belief. The latter—some masking mechanism which makes the M -conditions become non-conducive for P -investigation—was illustrated by the Computer case where the survey program was damaged whenever it surveyed a malfunctioning computer. In both scenarios a trustworthy epistemic method was implemented. The epistemic agent appears thus to be warranted in holding the relevant belief; and yet the Tracking Conditionals are not satisfied.

In the light of the principle of Variable Strictness and the principle of Compositionality Strictness the masking counterexamples to the DN theory do, though, constitute an exact analogy to the two mimicking counterexamples just discussed: in one case, the Computer case, the masking mechanism and the agent

work together and make up a 'closed system' with the overall disposition to believe P, that the computer is not malfunctioning when not-P is in fact the case; and in the other sort of scenario, exemplified by the Reversed Light case, it is the world and the masking mechanism that work together and make up a closed system with the overall disposition to make the agent form fallacious beliefs. Let us discuss this latter scenario first where the masking mechanism and the world make up a closed system with the overall disposition to make the agent form fallacious beliefs. Like the analogous mimicking scenario, this disposition of the world does not interfere with the epistemic disposition of the agent. In the mimicking case we said that a mimicking mechanism situated in the world did nothing by way of establishing a capability to get things right which the agent did not already possess. Likewise, in the present case, we say that the disposition of the world to make an agent form fallacious beliefs does nothing by way of depriving an agent of an epistemic disposition to form veracious beliefs which she possessed beforehand. The Reversed Light case constitutes thus a genuine counterexample to the DN theory: the agent is disposed to get things right regarding P although the Tracker conditionals are not satisfied. In the Computer case, on the other hand, where the masking mechanism and the agent make up a closed system with the overall disposition to form fallacious beliefs, it is now more natural to say that the epistemic method in question (running the computer's own survey system) does not dispose the agent to form true beliefs as to whether the computer is malfunctioning. A 'part' of the agent-computer system *is* so disposed, namely the part consisting of agent and computer except the masking mechanism. But, as our discussion of the principle of Compositionality Strictness brings out, this part possessing a certain disposition does not imply that the entire agent/method system possesses the same disposition. In the Computer case the DN theory thus gets things right after all: the agent is not disposed to form true P-beliefs.

5.2 A Dispositional Theory of Knowledge and Warranted Beliefs

The findings from Chapter IV have thus provided some progress. Some of the problem cases for the DN theory evaporate once sufficient attention is paid to the principle of Compositionality Strictness and the principle of Variable Strictness. But, more importantly, the discussion in the foregoing section has revealed an important common feature of the remaining problem cases: they all have the mimicking, masking or altering mechanism situated on the worldly side such that it functions independently of the enquiring agent. The principle of Variable Strictness makes it clear that the mimicking, masking or altering mechanism must function in a manner that secures some non-accidental relation between P and the agent's P-belief or, more specifically, between P and the agent's executing some specified P-enquiring method. For, if this relation is established in an accidental manner, specific for each particular scenario, then it will not have any impact on the semantic evaluation of the subjunctive conditionals featuring in the DN theory. This non-accidental relation between P and the P-probing can be established in either of two ways: i) the disturbing mechanism co-operates with P-matters to manipulate the agent's P-belief (the Holo case and the Computer case), or ii) the disturbing mechanism co-operates with the agent to manipulate P-matters (the Light case and the Reversed Light case). In short, the disturbing mechanism enters into a closed system with either P (i) or the P-probing (ii).

So much is, in fact, a mere consequence of the principle of Compositionality Strictness. But the discussion in the foregoing section has revealed that mimicking counterexamples to the DN theory moulded after ii) are merely apparent whereas those moulded after i) seem more steadfast and, correspondingly, that the masking counterexamples moulded after i) are merely apparent whereas those moulded after ii) seem more steadfast. And, interestingly, in the two issuing categories of genuine counterexamples the intuitive reaction is to emphasise that if the agent

came to believe such and such about P when executing the relevant epistemic method, then such and such would in fact be the case. Were Holo to believe P by watching a holographic representation chances are that P would not be the case. Holo is thus not warranted in believing P (despite the DN theory's verdict to the opposite) even if she in one particular setting regularly hits on the truth. And were Flip to believe the light is on in the Reversed Light scenario she would (again, contrary to the DN theory's verdict) be warranted in holding that belief. What we are focusing attention on here are, in fact, the *converse* subjunctive conditionals of those featuring in the DN Theory. Although the Tracking conditionals are satisfied in both Flip's and Holo's cases, an asymmetry emerges when considering the reversed subjunctive conditionals, which, for reasons which will become evident shortly, we will call *Tracker Conditionals*:

Tracker Conditionals

(DISPiii) $SB(\neg P)_{M\Box} \rightarrow \neg P$

(DISPiv) $SB(P)_{M\Box} \rightarrow P$

It is tempting, at this point, to object that no relevant difference exists between the two sets of subjunctive conditionals which could mandate the asymmetry considerations just outlined. Whenever the worldly states of affairs, P and $\neg P$, are such that they subjunctively imply that the agent is forming true beliefs regarding them under M-conditions, the agent's forming beliefs about those worldly state of affairs under M-conditions, must likewise subjunctively imply the veracity of those beliefs. That objection, however, would be premature. Recall the discussion in Chapter III of the dyadic nature of response-dispositionality. When considering an 'external' disposition of some object O to evoke a certain response R in subject S, there will always be an 'internal' counterpart, *viz.*, the disposition of S to produce R in response to O. A case in point was the external disposition of a coloured object to evoke a certain colour-sensation in S when observed and its internal counterpart, the disposition of S to enjoy a certain colour-sensation when

observing that object. Granted, usually there is a perfect co-extension between the internal and the external dispositional property. Usually, but not always as the Quirky case witnessed.¹⁴⁴ Here the external dispositional property of the object to appear yellow prevails while the corresponding internal dispositional property of Quirky to see the object as yellow is frustrated due to a ‘quirk’ in his neural wiring—Quirky would, very sadly, pass away before he could get a glimpse of his killer.¹⁴⁵ The reversed case, where the external disposition is obstructed by some altering mechanism but where the corresponding internal disposition remains intact was provided by the Chameleon case.

In the Quirky case the frustrating altering mechanism is located on the part of the subject. We therefore say that the subject and the altering mechanism make up a closed system with the overall disposition not to see killer-yellow things as yellow. Whenever this closed system is confronted with killer-yellow objects it will fail to recognise them. Likewise, in the Light case, the mimicking mechanism (remote control) is located in the method. We can think of Flip and her coin-tossing method as a closed system. Whenever they work together they possess the overall disposition to produce true beliefs. Considered as a system they are disposed to get things right. They work together in the same manner as the archer and his magic computer-controlled arrows work together: these make up a closed system with the overall disposition to hit the target. In Holo’s case, on the other

¹⁴⁴As introduced in Chapter III, section 3.7.

¹⁴⁵So, does Quirky possess the internal disposition of killer-yellow competence? According to the findings in Chapter IV, this is a case where part of Quirky has the disposition to recognise killer-yellow objects as killer-yellow, but where another part of Quirky, more specifically his quirk, is disposed to destroy the former disposition sooner than any manifestation would materialise. The ‘whole’ Quirky is, therefore, not disposed to see killer-yellow objects as killer-yellow (just as the whole circuit breaker is not disposed to conduct despite one of its central components being so disposed—*cf.* the Principle of Compositionality Strictness).

hand, the mimicking mechanism is located more to the worldly side. Here it is the P-state of affairs, the vase in the box, and the holographic apparatus which work together as a closed system. Whenever the vase and the holographic display thus work together they possess the overall disposition to *make* observers form true beliefs. They possess the disposition to be *tracked*. This compares to the second Zen scenario where the mimicking mechanism, the ‘arrow magnet’ is located in the target. Here the target and the mimicking magnet work together as a closed system with the overall disposition to *make* arrows hit the target. They too possess the disposition to be *tracked*.

If these considerations are correct, there are two and not just one disposition involved in the fine art of forming true beliefs. There is the disposition of the agent to track facts under M-conditions and there is the disposition of the world to be tracked under M-conditions. Two dispositions which by and large co-occur but which come apart in special cases, most notably cases where mimicking, masking and altering are involved. Hitherto tracking has been identified with the set of subjunctive conditionals characterising one of these dispositions, and, unfortunately, with the wrong one, *viz.*, the disposition of the P-state of affairs to be tracked by S. In order to avoid confusion we will continue to refer to these as *Tracking Conditionals*:

Tracking Conditionals

(Niii) $\neg P_M \Box \rightarrow \neg SB(P)$

(Niv) $P_M \Box \rightarrow SB(P)$

On the basis of the discussion of internal vs. external dispositionality in Chapter III we can see that the Tracking Conditionals characterise the external dispositional property of the world to be tracked—equivalent to the external disposition of some object to evoke a certain colour-sensation in S. The internal counterpart, the agent’s dispositional capability to track under M-conditions, should be stated according to the format for colour-competence—the internal

disposition of S to enjoy a certain colour sensation when observing the object—
i.e. as:

Tracker Conditionals

(DISPiii) $SB(\neg P)_{M\Box} \rightarrow \neg P$

(DISPiv) $SB(P)_{M\Box} \rightarrow P$

This is the reason why we so urgently needed the suggested alteration in our terminology. The classical Tracking Conditionals assign a property to the *world viz.*, the dispositional property of being co-operative for P-investigation under M-conditions. The reversed Tracking Conditionals, on the other hand, characterise the reverse of Tracking Conditionals—they assign a property to the epistemic *agent viz.*, the dispositional property of the agent—the tracker—to form true P-beliefs under M-conditions. We will stick to the tracking terminology but adopt Tracker Conditionals as the set of conditionals characterising the agent's dispositionality to form true beliefs. Likewise it will be useful to reserve the term *Warrant* for the kind of warrants an agent possesses when she satisfies the Tracker Conditionals.

With this finding in place we are now in a position to state an improved version of a dispositional theory which accounts for the remaining counterexamples. In order to demarcate it from its cousin, the DN theory, we will call it the Dispositional theory.

The Dispositional Theory

An agent S knows P iff

(DISPi) P

(DISPii) SB(P) based on execution of the epistemic method M

(DISPiii) $SB(\neg P)_{M\Box} \rightarrow \neg P$

(DISPiv) $SB(P)_{M\Box} \rightarrow P$

This formulation raises some issues that need to be discussed. Firstly, the alert reader will no doubt be puzzled about the antecedent in (DISPiii). A mere reversal of the (Niii) conditional gives $\neg SB(P) \text{ }_M \Box \rightarrow \neg P$; so why this strengthening of the antecedent to $SB(\neg P) \text{ }_M \Box \rightarrow \neg P$? Well, recall that $\neg SB(P)$ is an abridgement of the disjunction $SB(\neg P) \vee \neg[SB(\neg P) \vee SB(P)]$, that is, for the claim that S either believes P is false or believes nothing at all regarding P. This modest formulation was needed for (Niii) to be plausible since execution of the method M may be indeterminate—that is correct execution of M does sometimes leave the epistemic agent agnostic about P such as it for instance was the case in the DNA case in Chapter III, section 3.5. But, contrarily, the same modest formulation is entirely implausible as an antecedent in (DISPiii). Say M is characterised by a high degree of indeterminacy, that its correct execution generally does not issue in a verdict regarding P. In that case $\neg SB(P) \text{ }_M \Box \rightarrow \neg P$ could be trivially false since its antecedent could be vacuously true. Were it not the case that S would believe P under M-condition, this would be due to the second disjunct above: that S remains agnostic about P and that fact, on its own, does not at all indicate whether P holds or not. A measure for the agent's dispositionality to form true beliefs under M-conditions is only accessible when she as a matter of fact does form a belief about P, that is when the modest formulation is satisfied due to the first disjunct, that S believes P is false. Or, put differently, the agent's disposition to form true beliefs is only triggered when she actually comes up with a P-belief (just like the archer's disposition to hit the target is triggered by actually dispatching an arrow towards the target—not merely by entering the archery hall). We could of course, as it was indeed suggested in Chapter III, section 3.5, adjust the notion of M-condition to imply that the epistemic method M not merely is being exercised but, furthermore, has issued in some conviction regarding P. In that case we could stick to the plain reversion of the (Niii) conditional since $\neg SB(P) \text{ }_M \Box \rightarrow \neg P$ and $SB(\neg P) \text{ }_M \Box \rightarrow \neg P$, thereby, would be rendered equivalent.

Secondly, it might be questioned whether both (DISP_i) and (DISP_{ii}) are needed in the Dispositional theory. Given classical semantics, subjunctive conditionals with true antecedent and false consequent are false. So, given (DISP_{ii}) and (DISP_{iv}), (DISP_i) is trivially true; just as (DISP_{ii}) is trivially true given (DISP_i) and (DISP_{iii}). Likewise traditional semantics has it that subjunctive conditionals with true antecedent and true consequent are true. Thus, given (DISP_i) and (DISP_{ii}), (DISP_{iv}) is trivially true. The reply to this misgiving is simple: classical semantics is not correct. Nozick and others have already supplied cogent arguments for the rejection of the claim that subjunctive conditionals with true antecedent and true consequent are true and, as we saw in Chapter IV, it is a mere consequence of giving up this claim that we also must give up the claim that subjunctive conditionals with true antecedent and false consequent are false. One can thus be Warranted in believing P ((DISP_{ii}), (DISP_{iii}) and (DISP_{iv})) despite the falsehood of P \neg (DISP_i) just as one can hold a veracious P-belief ((DISP_i) and (DISP_{ii})) without being Warranted ((DISP_{iii}) and (DISP_{iv})). The Dispositional theory thus provides an independent account of Warrantability distinct from that of knowledge.

Thirdly, it may be questioned whether satisfaction of the Tracker conditionals (DISP_{iii}) and (DISP_{iv}) is sufficient to bring about a Warrant—in particular it may be questioned whether it ought to be supplemented with satisfaction of the Tracking conditionals (N_{iii}) and (N_{iv}) in order to bring about a genuine Warrant. The idea would thus be that only satisfaction of *both* Tracker conditionals and Tracking conditionals will count as necessary and sufficient condition for Warrantability. But this is not so. As for instance the Reversed Light case brings out the agent may be Warranted although the Tracking conditionals are not satisfied. Satisfaction of the Tracking conditionals is not even a necessary condition for knowledge. Granted, the more co-operative the world is in the epistemic project the better are the chances that the agent's effort will be rewarded with knowledge. But that is not yet to claim that a competent attempt, when only occasionally rewarded with a true belief, does not count as knowledge. This we

can illustrate with the archer analogy. An archer enters an archery hall where the targets are anti-arrow magnetic. Such an archer might well be competent (be Warranted) although he hardly ever hits the target. And if he, occasionally, performs a direct hit this ought to be considered a hit resulting from skill (knowledge). An epistemic analogue may be found in the Chameleon example: the brown-grey chameleon is sitting in the dark on a green leaf. Were an agent to observe it, and thereby bring about light, she would see it as green rather than as grey-brown (due to the creature's camouflaging skills). Here the world is not particularly co-operative for the epistemic project. And the Tracking conditionals are certainly not satisfied. Were the beast to be grey-brown she would not come to believe it were. Nevertheless, if her epistemic competence—her disposition to form true beliefs—on one particular occasion were to issue in a veracious belief that the beast is grey-brown (the chameleon might be a bit slow to swap colour on that particular day or what have you) then that belief would count as knowledge.

5.3 Tracker Conditionals versus Tracking Conditionals — A Dispositional Diagnosis

We said the Tracking Conditionals do not characterise Warrantability at all and that all the work in that respect is done by the Tracker conditionals. This claim is actually a little disingenuous. Due to the high degree of co-extensionality between the Tracker Conditionals and the Tracking Conditionals, the latter are as a matter of fact a very good approximation to the real thing.¹⁴⁶ In fact, as already observed, they only come apart from the real thing when mimicking, masking and altering powers are operating. A special case thereof accrues when both Tracking

¹⁴⁶If they were not, how could Dretske, Nozick, and with them many others have considered the Tracking Conditionals as, if not *the* correct, then at least the *best* candidate to opt for when providing a theoretical foundation for epistemic notions?

Conditionals *and* Tracker Conditionals are obstructed: assume some (surface green) object has a masked disposition to appear green, a disposition masked by rays emitted from the centre of the object with the effect that it as a matter of fact appears as blue.¹⁴⁷ Assume furthermore that Grue has a masked disposition to see apparently blue objects as blue, a disposition masked by a certain drug Grue is taking to the effect that these objects as a matter of fact appear green to him. Now, by a double-negation effect, Grue would see that surface green object as green whenever he looks at it. This is so, despite the fact that neither the relevant Tracker Conditionals nor the relevant Tracking Conditionals are satisfied. A similar case was among the problem cases for the tracking theory discussed in Chapter II. Moody, due to a minor brain damage, has erroneously come to believe two twins can be distinguished by a mole on the chin of the oldest twin (in this case Judy). Simultaneously with Moody's brain injury, Judy happens to develop a mole on her chin. You see where things are going: neither Tracker Conditionals nor Tracking Conditionals are satisfied. But due to the same 'double-negation effect', Moody would believe Judy were before him if she were and he would not believe Judy were before him if she were not.¹⁴⁸ A clear differentiation of the two dispositions involved provides the appropriate means to solve these puzzles. Focusing separate attention on Tracker Conditionals and Tracking Conditionals in turn, it becomes obvious that neither are satisfied.¹⁴⁹ The double-negation effect is only damaging when the two sets of conditionals are being intermingled.

¹⁴⁷This example is due to Johnston (1993).

¹⁴⁸Nozick's way of handling this case is, as already pointed out in Chapter II, highly unsatisfactory.

¹⁴⁹In order to prevent this sort of 'double-negation' counterexamples to the Dispositional Theory, it would be necessary to add a clause, (Dv), stating that the satisfaction of (Diii) and (Div) is not courtesy to the world's disposition not to make an agent form veracious beliefs under M-conditions. However, we have chosen not to add such a clause. For two reasons: i) double negation counterexamples are exceptionally rare and so without any practical significance for the theory,

Yet another problem for the DN theory was posed by the Jesse case. Here Stance happens to get a glimpse of the escaping Jesse's face, but only because his (Jesse's) mask slips just as he passes Stance. Intuitively Stance knows that the identity of the escapee is Jesse James (P). And yet (Niv), $P \text{ }_M \square \rightarrow \text{SB}(P)$, is not satisfied. This case illustrates very clearly the deficiency of the DN Theory. The reason (Niv) is not satisfied here is obviously that the world is not particularly co-operative for P-investigation. Had Stance not employed his method, *viz.*, perceiving, at that particular time and place the world would not have rewarded Stance for his epistemic effort; and generally the world would let P-investigators down.¹⁵⁰ But, surely, this non-co-operative feature of the world should not reflect badly on Stance and the status of his epistemic accomplishments. Stance has carried out his investigation irreproachably and his P-belief is, accordingly, a product of his dispositional capability to form true beliefs. The Dispositionality Theory provides for precisely that diagnosis: the Tracking Conditionals ((Niii) and (Niv)) are undermined (the world is not disposed to co-operate in the

and ii) our main task is not 'to get things right' anyhow—rather the project here is to provide a philosophical diagnosis of what is going on in these double-negation cases.

¹⁵⁰In general the world is non-co-operative in the epistemic project when its disposition to make an agent form true beliefs under M-conditions is masked, mimicked or altered. With the findings in Chapter IV (that there is no genuine masking mimicking and altering) it gets increasingly problematic to define non-co-operativeness the way we just did. At this point it would be more appropriate to define non-co-operativeness as an overall disposition of the world not to make the agent form true beliefs under M-conditions (dispositions to make the agent form fallacious beliefs under M-conditions being a significant, but not exhaustive, subclass).

Many great philosophers have pointed out how difficult it is to get to know oneself—'oneself' thus being a fraction of the world particularly disposed not to make the agent form veracious beliefs when inquiring (and here it is a pity we cannot any longer apply the terms masking and mimicking).

investigation) but the Tracker Conditionals ((DISPiii) and (DISPiv)) are satisfied (the agent is disposed to get things right nevertheless). By (DISPiii) were Stance to believe it was Jesse, it would be because he got a good glimpse of Jesse's face, and, needless to say, it would be Jesse passing by. By (DISPiv) were Stance on the other hand to believe it was not Jesse, it would be because he had a good glance of a mask-less somebody else—and the escapee would have been somebody else.

The Dispositional Theory allows us to locate disadvantageous epistemic factors more precisely than the DN Theory. In particular the Dispositional Theory makes it transparent whether it is the cognitive agent or the world (or both) that is blameworthy when something goes awry in the epistemic project. It is this flexibility that makes it capable of dealing satisfactorily with the following examples. Consider this sophisticated twist to the Jesse case: Jesse is not wearing the usual black bank robber's mask but rather an authentic Billy the Kid disguise. Were Stance to see Jesse wearing this disguise mask, Stance would not remain agnostic about the escapee's identity but actually come to believe something false, *viz.*, that the identity of the escapee is Billy (Q). Yet intuitions have it that Stance still knows P if he happens to get a good glimpse of Jesse just as the Billy disguise slips.¹⁵¹ Compare this with a scenario where Jesse is not wearing a mask at all but where Stance unbeknownst to himself has been given a Kid drug with the effect that he comes to believe of everyone passing by on a horse that it is Billy the Kid. This Kid drug is efficient but not impeccable—and (as you would expect) it 'slips' just as Jesse passes by on horseback and Stance comes thus, veraciously, to believe P. The two scenarios are identical but for the fact that the mask that slips at the crucial moment is external in the former case but internal in the latter case. And it is precisely this feature that explains why most of us would be more hesitant to ascribe knowledge to Stance in the latter case. The Dispositional Theory has the resources to articulate precisely what this demarcating feature

¹⁵¹The twist is due to Crispin Wright (1998c).

amounts to: in the former case it is the Tracking Conditionals that are compromised, whereas it is the Tracker Conditionals that are undermined in the latter case. In the latter case Stance, with his cognitive apparatus disabled by the Billy drug, is not disposed to get things right regarding P. In the former case he is.

The former scenario is very similar to one borrowed from Castle Land where Lucky Luke comes veraciously to believe there is a barn before him (P). But luckily so, for all the other barns in the area are covered behind papier mâché castle facades.¹⁵² Still Luke definitely knows P in this case. He has the barn right in front of him under ideal observation conditions and there is no deceptive trickery involved barn-wise. And so is the verdict of the Dispositional theory: were Luke to believe P, P would be the case; and were he to believe \neg P it would be because he believed there was a tree, field, horse or whatever before him and so \neg P.¹⁵³ The Tracker Conditionals are thus satisfied. The Tracking Conditionals, however are undermined. Were there to be a barn before Luke, he would not come to believe so.¹⁵⁴ Compare this with a scenario in which Luke stands before the only real castle in the area. Does he know there is a castle before him (Q)? No, this time he is a little too lucky.¹⁵⁵ Here the situation has been reversed: the Tracking Conditionals are satisfied but the Tracker Conditionals are not. Were

¹⁵²Freely improvised after Goldman's Barn Land case. See Goldman 1973.

¹⁵³See footnote 33*.

¹⁵⁴This case makes it clear how "were there to be a barn" and "were Luke to believe there were a barn" evoke distinct hypothetical scenarios. Were there to be a barn it might very easily be one of the disguised ones, in which case Luke would not come to believe there was one. Were Luke, on the other hand, to believe there was not a barn it would be, as mentioned, because he found himself in slightly different surroundings than the present one—surroundings of trees, fields, horses or, in short, barn-less surroundings. It is less likely that *these* alternate panoramas should comprise a fake castle—a barn surrounding.

¹⁵⁵Or so people tend to think—this is just a castle version of the classical fake-barn scenario.

Luke to believe Q, P would be the case. Intuitively, we would say Luke is not suitably equipped cognitively for Q-investigations in Castle Land. He is completely in the dark regarding a crucial feature of the epistemic settings for Q-investigations *viz.*, that almost all Q-appearances are fake. This is the reason he does not know Q in the actual case. And also the reason why he is not disposed to form a true belief regarding Q.¹⁵⁶ P-investigations, on the other hand, he is fully competent to carry out. The world will, of course, do its best not to reward his efforts, that is, it will do its best to hide the P-facts. However, if Luke overcomes these worldly difficulties and eventually forms a P-belief, this will be the manifestation of a disposition to form veracious beliefs regarding P.

In all these cases the Dispositionality Theory proves its worth and provides not only the correct verdicts, but also a diagnosis of the epistemological asymmetry featuring in *prima facie* symmetrical cases.

¹⁵⁶Note that the DN Theory gets things wrong here. It too points out that intuitions have it that Luke does not know Q in this case—but claims that precisely this is the outcome of the DN Theory since (Niii) is falsified: were it the case there were not a castle, Luke would still come to believe there was one. Because, the DN Theory continues, were there not to be a castle there would be a fake castle. For the reasons given in fn. *, this is a most implausible assumption. Were there not to be a castle it would be because the surroundings would be slightly different from the present ones i.e. the surroundings would be tree-, field-, horse-, or whatever-surroundings. Why assume slightly altered surroundings are fake-castle surroundings? If the answer is: in order to make as minor changes as possible in Luke's cognitive state of mind—the thought being that minor mindly alterations are justified on pain of larger geographical alterations—one is really advertising (Div) as an alternative to (Niii) (“were S to believe P” rather than “were \neg P to be the case”). In that sense, a good thought. *One* step towards the Dispositional Theory.

5.4 *Certainly Disposed*

The most significant result to be extracted from the Dispositional Theory, however, emerges when applied to the so-called background beliefs that have been a recurrent theme throughout. What is the epistemic status of these beliefs? Notoriously, they fail to satisfy the Tracking Conditionals. But what is their standing when evaluated relative to the Tracker Conditionals? Take the familiar

(B) I am not a brain in a vat

It is a familiar point that the DN Theory denies us knowledge of B since (Niii) is not satisfied: were $\neg B$ to be the case, I (or somebody else first-person singular) would still believe B. But that is only to emphasise that the world is not particularly generous when it comes to its supply of B-clues, that is, visible signs indicating whether or not B is the case. The only finding we can infer from that is that *if* the method involved in forming a B-belief is empirical in nature, for instance a belief formed by examining carefully whether the present surroundings are reminiscent of B-surroundings, *then* the issuing B-beliefs would not be particularly reliable—much less a manifestation of some disposition to form true B-beliefs. But the point is precisely that the method involved in forming B-beliefs is very different in nature. It is by no means an easy task to define the method involved in any detail,¹⁵⁷ but let it suffice to mention that it essentially involves *a priori* cognitive processes. And the fact that the method involved in forming B-beliefs is *a priori* does not automatically imply that its proper execution cannot dispose an epistemic agent to form true beliefs regarding B. Our mathematical organ (whatever *its* exact nature) is a paradigm example of a cognitive processor

¹⁵⁷The best bet might be “I can’t help believing” (Wittgenstein 1979, §277): a capability “that the human being acquires by means of observation and instruction. I intentionally do not say ‘learns’” (§279).

that disposes us to form true beliefs. Why not conceive of our organ for background beliefs in a similar manner?¹⁵⁸ In any case, once it is recognised that the method involved is (partly) *a priori*, the Dispositionality Theory only supports the idea that we are disposed to form true beliefs regarding B—the Tracker Conditionals (DISPiii) and (DISPiv) *are* satisfied. Were someone to believe she were not a brain in a vat she would not be (DISPiii).¹⁵⁹ How about the second Tracker Conditional (DISPiv)? Were someone to believe she was a brain in a vat, would she be one? One could of course object here that someone, call him Mental, has come to believe he is a brain in a vat due to a mental disturbance. If Mental were to believe he were a brain in a vat, he would not be one and (DISPiv) seems thus to be defeated. Such an objection would, however, only support the Dispositional Theory: Mental does not, intuitively, know that he is not a brain in a vat; the Dispositional Theory thus gets things right regarding Mental. What matters is not whether someone in particular would be a brain in a vat if he were to believe he was (just like it does not matter whether someone in particular who just happened to have had some haggis recently would die if he ingested a poisonous pill). What matters is whether *someone*, cognitively lucid and representative epistemic agent would be a brain in a vat if she were to believe she were. You, for instance, would you be a brain in a vat if you were to believe you were? Two scenarios spring to mind in which you would believe you were a brain a vat: (a) a scenario in which you have suffered some kind of mental disturbance which destroys you disposition to get things right regarding your bodily status; and (b) a scenario in which you have become victimised by a benign (?) scientist consciously letting you know the truth, that you really are a brain in a vat. Now,

¹⁵⁸Something along these lines appears to be the main thought in Ch.. 6 of Plantinga's *Warrant and Proper Functioning* (1993b).

¹⁵⁹ This is something she believes in all nearby worlds. All nearby worlds are also not-BIV worlds. Hence, the closest worlds in which she believes she is not a BIV are worlds in which she is not.

which of these two scenarios is most far-fetched? Admittedly, both scenarios are rather distant and thus it may be hard to demarcate which one is most distant. But personally I find scenario (a) most distant.¹⁶⁰ Whatever your verdict is, try evaluate the following B-beliefs relative to the Tracker Conditionals:

- M There are other minds.
- U The world is governed by uniformity.
- P There is a past.

Here it is even more obvious that the Dispositional theory grants one Warrants. Were someone for instance to believe M, M would be the case. And were someone to believe $\neg M$, M would not be the case. Again, one could come to believe $\neg M$ due to a mental disturbance in an M-scenario. But if one (you) were to believe $\neg M$, chances are that this would be due to the absence of other minds in your surroundings.

It is tempting to object here, that if $\neg M$ and some smart trick scenario has been set up to pretend M, an arrangement including humanlike automata, say, then (DISPiii) would not be satisfied. Let us call realisation of such a trick scenario T_M . In that case, were someone to believe that M in a T_M -scenario, $\neg M$ might well be the case and (DISPiii) is thus not satisfied. This objection is reminiscent of a misgiving considered in Chapter IV: were Fred to ingest poison and also to ingest some antidote, the poison would not cause illness or death. And the reply is similar too: all that can be established by the example is that an epistemic agent is not disposed to form veracious beliefs when it come to the belief (M and $\neg T_M$).¹⁶¹

¹⁶⁰Exercise: conceive what scenario (a) would be like—*believing* that you are a brain in a vat. Is the scenario that strikes you one in which you still are embodied?

¹⁶¹ And it is dubious whether the example shows that much. Everything here hinges on the first Tracker conditional: $SB\neg(M \text{ and } \neg T_M) \square \rightarrow \neg(M \text{ and } \neg T_M)$. As in the BIV case the truth-value of

And, to be sure, this does not imply that the agent is not disposed to form veracious beliefs about M.¹⁶² According to the Dispositional Theory one can thus know that M and yet fail to know that (M and $\neg T_M$).

The usual objection at this stage is to point out that M implies $\neg T_M$ and that knowledge of M thus, given Closure, ensures knowledge also of $\neg T_M$. But given one knows M and also knows $\neg T_M$ it appears highly suspicious to claim that one does not know (M and $\neg T_M$). There is a very efficient, but largely ignored, response to this line of objection: namely that M does not imply $\neg T_M$! If one envisages the T_M -scenario as one involving artificially designed robots, and thus (intentional) agency (robot design), it might even be argued that it is $\neg M$ which implies $\neg T_M$. But in any case the two facts M and $\neg T_M$ are mutually independent. An environment inhabited with minds may and may not in addition involve a set-up that deceives the epistemic agent to believe there are other minds where no minds exist. Likewise, a scenario with such a seductive mechanism may or may not, in addition to figments of imagination, be inhabited with real minds.

It is interesting to run the same sort of reasoning for the BIV case. The core issue here is whether or not an agent is Warranted¹⁶³ in believing W, that there is an external world. And it is rather obvious that an agent is Warranted in believing W. Both (DISPiii) and (DISPiv) are satisfied: were an agent to believe there is an external world there would be one, and, equally indubitable (if not more so) if an

that conditional depends on which $SB\neg(M \text{ and } \neg T_M)$ scenario is 'closest': one in which S believes $\neg M$ (due to an environment void of minds) or one in which S believes T_M (because she has detected T_M) or, finally, one in which she simply believes both $\neg M$ and T_M (due to a mental disturbance).

¹⁶² Due to the principle of Variable Strictness. Or, almost so. The Principle of Variable Strictness says that $A \Box \rightarrow C$ does not imply $(A\&B) \Box \rightarrow C$. What we adhere to here is the principle that $B(A) \Box \rightarrow C$ does not imply $B(A\&B) \Box \rightarrow C$.

¹⁶³ 'Warranted' as defined in section 5.2 as satisfaction of (DISPii), (DISPiii) and (DISPiv).

agent were to believe that there is no external world, there wouldn't be an external world. As long as the epistemic project is defined merely as W-investigation the world is extremely helpful in that project and, even, disposed to make the agent form veracious beliefs regarding W. The difficulty accrues only when we consider that epistemic project (W-investigation) carried out in a fraud environment, an environment involving some seductive mechanism (such as computer-controlled stimulation of the nerve-endings) which manipulates the agent to believe things about the external world completely independently of the real state of affairs and, indeed, independently of the existence of the external world. Call realisation of such a mechanism T_W . The question then is whether an agent is Warranted in believing (W and T_W) or, equivalently, Warranted in believing that she is not a BIV.¹⁶⁴ As already mentioned, the answer here hinges on what we take to be the 'closest' scenario in which the agent believes (W and T_W)¹⁶⁵ (that she is a BIV). Is

¹⁶⁴That is, a BIV from a Putnam BIV-scenario where the external world has been blown up and only the BIV society in the vat remain. The mere belief that one is not a BIV is: $\Box T_W$.

¹⁶⁵Interestingly, believing ($\neg W$ and T_W) has a smell of Moorean paradox to it (I believe P but $\neg P$): if one believes ($\neg W$ and T_W) one believes $\neg W$ and one believes T_W . That is, one believes there is no external world and, simultaneously, that one is being tricked to believe there is an external world. But that is contradictory: one cannot believe one is being tricked to believe there is an external world if one does not believe there is an external world! Hence, one cannot consistently believe one is Putnam-BIV. One can only believe one is a being tricked to believe there is an external world (that one is a BIV) if one believes there is an external world. Here is an analogy: one cannot both believe that one's wife is not faithful and that one's wife deludes one to believe she is. If one believes she is not faithful one can only believe she *tries* (unsuccessfully) to delude one into believing she is. Likewise, given one believes there is no external world one can only believe some seductive mechanism tries, unsuccessfully to make one believe there is an external world. One can have a *suspicion* that there might not be an external world and simultaneously

the closest SB(W and T_w))-scenario one where an embodied S is suffering some mental disturbance or is it one in which a BIV- S has been informed about her disembodied status by an honest scientist? Only in the former case is (DISPiii) not satisfied. But in any case it must be emphasised that an agent may be Warranted in W , according to the Dispositional Theory, although debarred from Warrants for (W and T_w).

Likewise the Dretskean contrast consequences are knowable according to the Dispositional Theory. Recall Stripe from the Zebra case (Chapter II, section 2.1). Does he know the animals before him really are zebras and not cleverly disguised mules? Yes: if he were to believe the animals were not cleverly disguised mules, they would not be; and if he were to believe they were, it would be because he had some inside information to that effect. So, unless Stripe is a very naive person by whom people amuse themselves by making him believe all sorts of odd things, chances are that his inside information is veracious and that the animals really are disguised. Close to this version is the scenario in which Stripe simply suffers from epistemic paranoia¹⁶⁶, and comes to believe such things as the ‘zebras’ probably are disguised mules merely by entertaining that possibility (e.g. due to an inquiring son) and thus fails to satisfy (DISPiii). But in this case Stripe is not disposed to form true beliefs regarding that kind of proposition and we would not ascribe knowledge to him—not even if he happened to be right on this particular occasion. Yet a third possibility is that Stripe comes from a poor country where zoo fakes are normal whenever that facilitates financial savings. He visits London Zoo shortly after his arrival in England and, given his past, has good reason to believe that the zebras really are mere mule-substitutes. Still, although

believe that one is being tricked to believe there is an external world. But, as noted in Chapter II, what counts is the *overall* belief of the agent.

¹⁶⁶‘Epistemic paranoia’ denoting a pathetic urge to pay *serious* attention to all sorts of far-fetched error-scenarios.

mentally lucid, Stripe is still not Warranted in his unusual belief.¹⁶⁷ He is, like Lucky Luke in front of the (only real) castle in Castle Land, ineptly equipped cognitively for the kind of investigation in question. He is completely in the dark regarding a crucial feature of the epistemic settings for zoo investigations. Likewise, we can say that Lucky Luke, had he, without any good reasons, formed the eccentric belief in front of one of the fake castles that it properly was a fake castle with a barn behind, he would be in the same position as (epistemically paranoid) Stripe: his belief, although true, would not be Warranted since moulded in a mind suffering from epistemic paranoia.

It might, quite reasonably, be objected that no substantial account has yet been provided of *what* the distinction between epistemic paranoia and reasonable doubt consists in; an account, that is, exceeding a socio-descriptive recording of where this line actually is drawn (in *our* linguistic community). All that has been established is that epistemic support is gathered on an invisible¹⁶⁸ tray of background information and that no further evidence provides further guidelines for identification of that tray—hence its invisibility.¹⁶⁹ With what right do we blame someone who does not grasp it? Our only justification for doing so is to appeal to the fact that some epistemic methods might be (partly) *a priori* and that this fact in itself does not necessarily degrade their epistemic status. We are blessed with an organ, an inner eye, for which the invisible tray has become visible.¹⁷⁰ Why should activation of this organ, auto-introspection, not dispose us

¹⁶⁷Although warranted: he has, after all, good reasons for believing the creatures are fakes. But the term Warrant is, recall, reserved for a warrant that satisfies the Tracker Conditionals.

¹⁶⁸Invisible in the sense that we only very rarely pay attention to them. See the discussion in Chapter II, section 2.1.

¹⁶⁹Cf. the discussion in Chapter II pp. *.

¹⁷⁰If you do not believe in the existence of such an organ, reflect what it would be like to lack it. One would either come to believe background beliefs on a purely random basis—or not at all. It

to form true beliefs? And why should we not say that someone with a malfunctioning inner eye is as epistemically disabled as someone with poor eyesight? This, however, is only to claim that we *may* be Warranted in believing all those things we ordinarily take ourselves to be Warranted in believing. The Dispositional Theory is externalistic: it provides an account of Warrantability and knowledge according to which we are in fact Warranted in holding a good majority of those beliefs that make up our complete worldview *if* the world, by and large, is designed as we think. And, again, if the world is so designed, we are Warranted in believing that too. But there is no way we will ever find out. In that sense Warrantability is an externalistic notion.

5.5 A Warrant for Tracking

We have said, and have now (at least to some degree) justified, that there are two distinct dispositions involved in the epistemic project. Two dispositions with a high degree of extensional overlap, but yet crucially different in nature. They are: the disposition of the world kindly to co-operate in the epistemic project by delivering veracious clues under M-conditions; and the disposition of the agent to form veracious beliefs under M-conditions. We have also said that the two dispositions are CA- analysable and thus precisely characterised by the Tracking Conditionals and Tracker Conditionals, respectively. Finally the claim was that the DN Theory has confused the two and focused attention on the wrong one—the Tracking Conditionals, whereas we have now seen that it is the Tracker Conditionals which are significant when evaluating the epistemic agent.

Admittedly, this sounds a little immodest. Furthermore, it does not pay sufficient attention to the epistemic role of the Tracking Conditionals in the overall picture.

would be like lacking the mathematical organ (and so being unable to spontaneously and sincerely believe e.g. $1+1=2$)—only much worse.

As a matter of fact both sets of conditionals are of crucial relevance in epistemology. True, the Tracker Conditionals are significant when we wish to evaluate the epistemic agent. But is it not equally important to evaluate the epistemic settings? Certainly it is if we reflect that there is a pragmatic aspect to epistemology: the enterprise of developing precise and visible signposts for competent *informants*—cognisers worth consulting when various pieces of information are needed for all kinds of everyday purposes.¹⁷¹ For the sake of this pragmatic project it is equally important to develop precise and visible signposts for trackable *states of affairs*—worldly states of affairs that are likely to reward one's epistemic efforts.¹⁷² One obvious reason this project is worth undertaking is that by using signposts for epistemically conducive settings, one is more likely to hit on competent cognisers when consulting the epistemic agents conducting inquiries in these worldly compartments. By contrast, if one wishes to learn something about barns it is a bad idea to look for experts in Castle Land. Matters are not that simple, though. Sometimes the untrackable nature of a particular worldly part is compensated by the potential gain for the epistemic inquiry and some highly qualified trackers might thus be found inquiring in an environment characterised by a low degree of trackability. Take theology. Or psychoanalysis. Many competent people have carried out extensive investigations in these areas

¹⁷¹See Craig 1990. Needless to say, no philosopher is particularly concerned with this pragmatic purpose. Yet, it might be constitutive for the pre-philosophical 'common-sense' intuitions governing the epistemic notions—the intuitions philosophers subsequently are struggling to provide a finer-grained theoretical foundation for. In that sense pragmatism might influence the 'raw material' for the philosophical investigation. (It might, of course, be that things work the other way round: that the epistemic notions, properly understood, only are realised as result of philosophical reflection (seeing the light) and that the mundane counterparts are mere shadow-like imitation of the real thing (Plato).

¹⁷²Thanks to Fraser MacBride for pointing this out.

despite the fact they are not particularly tracking conducive. And certainly this need not imply that the cognisers have not been Warranted at the end of their enquiries. Likewise certain areas of theoretical physics are characterised by an alarmingly low degree of trackability.

But even in these cases of a relative untrackable nature, the most competent trackers are those who are well acquainted with the conduciveness of their present environment, that is, are acquainted with the Tracking Conditionals, even when carrying out their project in a non-conducive environment. Indeed, to be radically mistaken about the Tracking Conditionals for one's epistemic settings will obscure the epistemic project¹⁷³. Even the most ingenious and talented physicist would be useless if she was not sensitive to the fact that the world is more cooperative for her epistemic project when investigated in the lab than in the wild. The regularities she is attempting to reveal provide veracious clues when isolated in the lab—but might well be masked by numerous intervening regularities in the wild. As this example illustrates, clear and visible signposts for conducive environments are not only useful when searching for appropriate circumstances in which to carry out one's investigations, but also might be useful in cultivating such circumstances. A full discussion of these matters, and in particular how they impinge on the evolution of science and their wider socio-psychological implications, would take us too far afield. They are mentioned here merely as a reminder that Tracking Conditionals still ought to play a crucial role in epistemology despite the fact that it is the Tracker Conditionals we ought to focus attention on when evaluating the epistemic *agent*—in particular, when evaluating whether the epistemic notions (knowledge and Warrant) can be ascribed to a particular agent.

¹⁷³As witnessed by for example Lucky Luke's ignorance of being inquiring in Barn Land.

5.6 Scandalous Scepticism

The discussion in the section ‘Certainly Disposed’, might lead to the conjecture that the Dispositional Theory complies with the principle of Closure. As we have seen, the classical counterexamples to Closure provided by DN Theory do not prevail in the Dispositional Theory. These counterexamples are the instances of Closure which usually feature in the standard set-up of the Sceptical Trilemma:

Sceptical Trilemma

- (A) We are warranted in believing some mundane empirical fact P (e.g. I have a hand)
- (B) We are not warranted in rejecting some typical sceptical hypothesis incompatible with P (e.g. I am a BIV)
- (C) Closure

Notoriously, the sceptic applies this Trilemma to support B and C on pain of A. Kant’s reaction to the issuing position is famous:

it must still remain a scandal to philosophy and to the general human reason to be obliged to assume, as an article of mere belief, the existence of things external to ourselves (from which, yet, we derive the whole material of cognition for the internal sense), and not to be able to oppose a satisfactory proof to any one who may call it in question. (1933, p. 34 n.)

as is Heidegger’s retort:

The ‘scandal of philosophy’ is not that this proof is yet to be given, but that such proofs are expected and attempted again and again. (1962, I.6)

The *real* scandal, however, is not that there is no satisfactory rebuff to the sceptic, *contra* Kant, nor that such a rebuff is both expected and attempted, *contra*

Heidegger. The scandal lies, rather, in the *way* this has been attempted.¹⁷⁴ The Sceptical Trilemma consists of two intuitively highly plausible claims, A and C, and one intuitively directly repugnant claim, *viz.*, B. Nevertheless the prevalent reaction to the Sceptical Trilemma in recent epistemology very often proceeds via a rejection of C (!) in favour of A (!!) and B (!!!)¹⁷⁵—or, worse, rejection of A (!!!!)¹⁷⁶, in context¹⁷⁷, combined with *neglecting* B (!!!!)¹⁷⁸ in other contexts.¹⁷⁹ Or, worse still, a combination of the two.¹⁸⁰ The greatest of all virtues of the Dispositional Theory is that it has a sane way of handling the Sceptical Trilemma, *viz.*, by supporting the intuitively plausible A and C and rejecting the repugnant B.

The Dispositional Theory supports Closure in the sort of applications made of it for sceptical purposes, the kind of Closure between a global sceptical hypothesis and some incompatible mundane empirical fact. And it does so by granting us warrants and knowledge both of the empirical fact *and* of the negation of the sceptical hypothesis, in tandem. However, when some *local* sceptical hypothesis is involved, the question of the validity of Closure becomes more precarious. Recall the Lucky Luke scenario from Castle Land in which Luke knows there is barn before him (P) if there is, but fails to know there is a castle before him (Q) even if there is one. This is so because Luke is inappropriately equipped, cognitively, when it comes to Q-investigations in that he is mistaken about a crucial Q-relative aspect of the settings (that most Q-appearances are fake). How are matters when it comes to \neg Q investigations? Well, on the face of it, it appears Luke does not know

¹⁷⁴Which maybe captures aspects of both Kant and Heidegger's misgivings.

¹⁷⁵Dretske 1970, 1971; Nozick 1981; McGinn 1984.

¹⁷⁶Hume 1955; Strawson 1985; Stroud 1984; Unger 1975.

¹⁷⁷Wittgenstein 1979; Quine 1969.

¹⁷⁸McDowell 1994, 1995.

¹⁷⁹Craig 1990; Dretske 1981; Williams 1996; Lewis 1996; Hetherington 1992.

¹⁸⁰DeRose 1995.

$\neg Q$ according to the Dispositional Theory since (DISPiii) does not hold: were Luke to believe $\neg\neg Q$, i.e. Q , it would be because he was attending to one of the fake castles and $\neg Q$ would, in fact, be the case. But since P implies $\neg Q$, and Luke knows P , Closure has it that Luke also ought to know $\neg Q$ if P . Is Closure not undermined locally then?

Indeed, it would be a serious, if not decisive, blow to the Dispositional Theory if it precluded Luke from knowing such matters as $\neg Q$. Not only would it, locally, undermine the highly intuitive principle of Closure, it would further yield an utterly implausible result, *viz.*, that Luke does not know there is no castle in front of him—even if standing in the middle of the most deserted part of Castle Land surrounded by only grass and bushes. Fortunately Luke does know $\neg Q$. We just have to be a little careful when describing how he comes to know this. All the Dispositional Theory asserts is that Luke cannot come to know $\neg Q$ by this particular method, *viz.*, perception. But a little reflection will reveal that perception, although a common way of getting to know Q , would be a highly unusual manner in which to come to know $\neg Q$. Let us assume you too know $\neg Q$ that there is no castle before you just now. How do you know? Well certainly not by observing some $\neg Q$ -fact, i.e. seeing a castle not being before you. Rather you look around (if you are very scrupulous) and see your desk, computer screen, bookshelves and (if you go to extremes) the empty lawn outside your window, and *thus* assure yourself you are sitting in your office without a castle before you. It is via a warrant for R , S and T states of affairs incompatible with Q that you gain your warrant for $\neg Q$. Likewise for Luke. He looks around and sees a deserted field, a few bushes and other R , S , T facts incompatible with Q and *thus* gains a warrant for $\neg Q$. The method involved, unnoticed as it might be, is *inference* from these R , S and T facts. And the Dispositional Theory gladly bestows knowledge of R , S and T on Luke. Knowledge that, via inference, issues in knowledge of

$\neg Q$.¹⁸¹ Notice that this is also the explanation (or diagnosis) of why Luke cannot gain a warrant for $\neg P$, that there is no barn before him, when positioned before one of the fake castles. The only way he can gain such a warrant for $\neg P$ is by way of a warrant for the incompatible Q and then infer $\neg P$ from Q and thus get his warrant for Q transmitted through to $\neg P$. But Luke never gains a warrant for Q —hence there is no warrant to transmit through to $\neg P$.

5.7 Transmission of Warrant

The principle of Transmission says:

Transmission

A particular warrant, W^* , for some proposition P is transmissible through to any (known) consequences Q of P . Thus for an epistemic agent to be warranted by W^* in believing P , and to recognise that Q is a consequence of P , is to be warranted by W^* in believing Q .¹⁸²

The principle of Transmission is thus stronger than the principle of Closure since the latter merely states that if an epistemic agent has some warrant for P , she also has some (but not necessarily the same) warrant for any (known) consequence of P , Q . At the same time Transmission is weaker than Closure in that it merely states an agent *can* come to be warranted in believing any consequence of a

¹⁸¹In general one never satisfies the antecedent in the first tracker conditional, $SB(P)_{\text{perception}} \square \rightarrow P$ when P is a negative existential claim: one never comes to believe such P under perception condition—one only comes to believe P under perception-and-inference conditions (how should it be possible to perceive nothing?). The first tracker conditional is thus trivially true (impossible antecedent) in the same manner as Nozick's second tracking conditional, $\neg P_{\text{intellectualisation}} \square \rightarrow \neg SB(P)$, is trivially true for mathematical P 's.

¹⁸²(Close enough to) Wright's formulation in Wright (1998b).

warranted belief. Not that she is as a matter of fact so warranted. Whether she is, hinges on whether she actually has carried out the relevant inference. You can come to be W^* -warranted in believing there is no castle before you by inferring it from your W^* -warranted belief that you are sitting in your office. But you might easily not be so warranted in believing there is no castle before you for the question, whether there is a castle before you, might never have crossed your mind. Other consequences of your warranted beliefs might have crossed your mind, but the inference in question might have been too complicated for you to carry out instantaneously. Fermat's last theorem is a consequence of known propositions—but it took us a while to get the knowledge transmitted.

Is Transmission universally valid? Are there any consequences of warranted Ps to which we cannot transmit the warrant in question? Consider the following inference:

- (P) Paul is in pain; therefore
 (Q) There are other minds

Suppose there is a particular warrant W^* available for our belief in P. Is that warrant transmissible to Q? Nozick's test for acquisition of knowledge¹⁸³ in this manner via inference is whether an agent's application of that method (inference) satisfies (Niii) and (Niv) (Nozick 1981, pp. 230-40). Whether an agent, call him Marquis, would not have believed Q by this method had $\neg Q$ been the case and would have believed Q had Q been the case. Nozick furthermore remarks that if Marquis would not have believed Q by using this method it would be because Marquis would not have believed P. Likewise, if Marquis were to believe Q by using this method that would be because Marquis would believe P. Bracketing

¹⁸³Nozick is exclusively concerned with knowledge, but there is no obvious reason why his arguments on this particular aspect should not be applicable to warrants too.

together the possibilities of (i) lack of cognitive capability to carry out the inference, and (ii) neglecting the consequences, this is a plausible idealisation. Nozick concludes from these considerations that acquisition of knowledge via inference is possible when the following revised conditionals are satisfied:

(Niii') $\neg Q_{I/P} \square \rightarrow \neg SB(P)^{184}$

(Niv') $Q_{I/P} \square \rightarrow SB(P)$

conditionals that generally fail when Q is the negation of some sceptical hypothesis (as our Q above). Nozick writes:

When, for some sceptical hypothesis SK, an inference is made to not-SK from another statement p which is tracked and known, this condition [Niii'] is not satisfied, and so one does not come to know that not-SK through that inference. Let q be the statement (not-SK) that S is not in the (relevant) tank on Alpha Centauri, and let P be the statement that S is sitting and reading in Massachusetts. S knows that p entails q , and infers q from p [...]. If q were false and he were floating in that tank, he still would believe p (and infer q from p). (*Ibid.*, p.231; my square brackets)

The reasoning is the same in the case of our example. Were Q to be false, were there no other minds but only an abundance of very clever automata, Marquis would still believe P (that Paul is in pain) and, crucially, believe this on the basis of the very same evidence, *viz.*, a certain 'behaviour'. Hence, in a Nozickean spirit, one might likewise suggest that Marquis could not transmit his warrant for

¹⁸⁴Subscript I/P is to be read as: 'the Method: inference from P, has been implemented'.

P to Q by inference since Marquis would have had that warrant for P even if Q were false.¹⁸⁵

The Dispositional Theory, however, would take these considerations to indicate only that the worlds co-operativity—captured by the *Tracking Conditionals*—is not always closed under implication: that the world sometimes co-operates and is helpfully disposed to deliver veracious clues about certain subject matters, P, without necessarily being disposed to deliver veracious clues about certain other subject matters, Q, although P and Q are logically associated. But nothing has so far been said to the effect that Marquis, if Warranted in believing P can become Warranted also in believing Q by the principle of Transmission. What is at stake here, the Dispositional Theory would insist, is whether Marquis is disposed, on the basis of W^* , to form veracious beliefs about certain things, Q, inferred from P, given that he, on the basis of W^* , is disposed to form true beliefs regarding P—that is, whether Q would be the case were his inference to issue in Q and $\neg Q$ would be the case were his inference to issue in $\neg Q$. Thus:

(DISPiii) $SB(\neg Q)_{/P} \square \rightarrow \neg Q$

(DISPiv) $SB(Q)_{/P} \square \rightarrow Q$

Both (DISPiii) and (DISPiv) must be satisfied for Q (that there are other minds), if they are so for P (that Paul is in pain).

It is important here to state precisely what the method consists in i.e. which method Marquis applies to gain a Warrant for P. Suppose this method is observation of Paul's behaviour and that execution of this method issues in a Warrant W_B for:

¹⁸⁵Bracketing the complications stated in Chapter III: that one, in a Nozickean spirit, *can* transmit the warrant to (Q & P) (which also is a necessary consequence of P). This is so since one would not have the warrant for P had (Q & P) been false.

(B) Paul is displaying pain behaviour

Given our discussion in the section ‘Certainly Disposed’, W_B , and any other empirical warrants for that matter, are useless on their own as a Warrant for Q since both Q and $\neg Q$ *per hypothesis* are completely indifferent to W_B . The world is not at all co-operative in delivering behavioural clues on Q-matters and all behavioural observation is therefore a poor indicator for detection of Q versus $\neg Q$ scenarios.

All this is fairly non-controversial. Indeed it is just the well known voice of the sceptic. What is more controversial is that W_B , on its own, is also a useless Warrant for P. Granted, *prima facie* P-matters are not indifferent to W_B in the same manner as Q-matters are. What then prompts the claim that W_B on its own is useless as warrant for P? It is important here to clarify what it means to be warranted merely by W_B . In fact it is very hard to imagine what it would be like to be warranted merely by W_B for, according to our usual way of conceiving of these matters, W_B always co-occurs with a background belief to the effect that

(A) Paul’s behaviour has a mindly aetiology, i.e. it is (partly) a reflection of some state of mind

a belief which in turn is Warranted by some Warrant W_A .¹⁸⁶

It is very hard to imagine what it would be like to possess a W_B without simultaneously possessing a W_A , the warrant supporting the associated background belief A. Even harder to imagine what it would be like to enjoy a B-

¹⁸⁶Or, at least we like to think we are so Warranted in believing (A). The questions whether we really are and what precisely such a W_A consists in are questions that lie beyond the scope of the present essay.

belief without simultaneously enjoying the associated A-belief. Nevertheless, it is a genuine possibility. And were this possibility to be instantiated, there would be no good reason to suppose that the belief about Paul's behaviour, on its own, would trigger as much as a belief about Paul having a mind—far less a mind which is in pain. Likewise it is a genuine possibility that the relevant background belief A is displayed, but completely unwarranted. In the absence of such a W_A , there is no good reason to suppose that W_B , on its own, should amount to a Warrant for the belief that Paul is in pain. When we say that P (Paul being in pain) is being supported by W_B in a manner that Q (there are other minds) is not, we really mean that P is being supported by W_B *in conjunction with* W_A . We have become accustomed to the idea of a W_A to such a degree that we never pay any attention to it. It is never explicitly mentioned, but always assumed as a tacit presupposition. And for all practical purposes, it needs no mentioning. Nevertheless, the fact remains that possession of W_B , all on its own, does nothing by way of disposing Marquis to form true beliefs about P.¹⁸⁷

The combined Warrant W_{A+B} consisting of W_A and W_B , on the other hand, would serve as Warrant for P. But *this* Warrant is also transmissible to Q. Marquis is disposed to form veracious beliefs regarding Q when he has W_{A+B} . As the discussion in 'Certainly Disposed' brings out, he is disposed to form true beliefs about Q as soon as he has W_A . *A fortiori*, he is so disposed under W_{A+B} conditions.¹⁸⁸ If, on the other hand, the sceptic were right and Marquis is not

¹⁸⁷ Although W_B combined with the relevant (unwarranted) background belief, would provide the agent *reasons* to believe B (good reasons?). W_B might thus serve as warrant for P although failing as Warrant.

¹⁸⁸ To be sure, nothing that has been said so far is likely to impress the sceptic. For, could it not be that the collateral belief, that Paul's behaviour has a particular aetiology, despite its certain phenomenology, is not Warranted after all? All that has been said is that *if* it is Warranted by W^* , that W^* is transmissible to Q.

disposed to form veracious Q-beliefs when he has W_A , i.e. if W_A does not serve as Warrant for Q-beliefs, then nor will W_{A+B} serve as Warrant for Q-beliefs. But in that case W_{A+B} will not serve as Warrant for P-beliefs either.

Transmission thus holds. *If* Marquis has a warrant W^* for believing P he can, by inference, come to be warranted, by that same warrant W^* , that Q. The principle of Transmission ought, however, to be distinguished from the principle of Acquisitionary Transmission.

Acquisitionary Transmission

One can *acquire* a Warrant for any consequence of P, Q, by transmitting one's Warrant for P to Q by inference and thus come to be Warranted in Q *for the first time*.

To illustrate what we have in mind, consider the following case: Truman has a Warrant W_A for believing:

(A) The cause of rain is condensed rain clouds

Assume furthermore Truman has a Warrant W_B —he is being rained on, say—for believing:

(B) It is raining

Truman now has a Warrant, W_{A+B} for believing:

(P) Condensed rain clouds are passing by

Since Transmission holds for Warrants, Truman can now by inference come to be Warranted in the consequence of P:

(Q) I'm not the star in a Truman Show (since there are real clouds around)¹⁸⁹

Note, however, that W_B , on its own, is useless as a Warrant for Q. Admittedly, W_B does transmit through to Q—to the extent it Warrants Truman in believing P (zero) it also Warrants Truman in believing Q (zero, again). But it is only given a W_A for the collateral A-belief¹⁹⁰ that W_B is any good for Truman as a Warrant for P and *a fortiori* as a Warrant for Q. Is there any difficulty here? Can Truman not just acquire the demanded Warrant W_{A+B} by inference from P? Given that he possesses a W_{A+B} Warrant for P he can, by inference, gain a W_{A+B} for Q (Transmission). But he cannot come to *acquire* W_{A+B} for Q by inference in this manner since Q is a necessary condition for A:

$A \rightarrow Q$

Assume now that Truman was not Warranted in believing Q before he carried out his inference (i.e. that he really *acquired* his W_{A+B} for Q by inference). How, then, could he be Warranted in believing A before he carried out his inference from P to Q? As we have seen Truman cannot be Warranted in believing P without also being Warranted in believing Q. Hence, Truman was not Warranted in believing P before he carried out the inference from P to Q. But if he was not Warranted in

¹⁸⁹Cf. *The Truman Show*—a movie which hero, Truman, is victimised in a 24-hour television show (with himself as a star!) which, unbeknownst to him, is recorded in an artificial town with a multitude of hidden cameras and where everything, including rainfall, is fake. The point of the show is to pursue a psychological experiment: is it possible to maintain the illusion, i.e. to maintain the seductive veil surrounding Truman—is it possible, for instance, to maintain the illusion that rainfall is caused by rainclouds.

¹⁹⁰And, again, a firm conviction that A is true will not do the trick—what is called for is a *Warrant* that A is true.

believing P before he carried out that inference, there was no Warrant for P to be transmitted through to Q by that inference. Hence, transmitting the Warrant for P through to Q is not an option for Truman for *acquiring* a Warrant for Q.

Likewise, Marquis cannot acquire a Warrant for his Q-belief, that there are minds, by inference from P, that Paul is in pain. If Marquis lacks such a Warrant for Q before he carries out this inference he also lacks a Warrant for A, that behaviour has a mindly aetiology (for the same reasons as those rehearsed in the Truman case). But if Marquis lacks a Warrant for A before he carries out his inference from P to Q, he also lacks a Warrant for P before he carries out his inference from P to Q, since W_B only serves as Warrant for P if associated with a Warrant for A. Hence, there would be no Warrant to transmit through from P to Q. So inference is not an option for Marquis either for acquiring a Warrant for his Q.

Where does this lead us? Is it a consequence of the Dispositional Theory that one, generally, never gains a Warrant by inference, for some proposition P unless one is already Warranted in P before one undertook that reasoning? That would certainly be catastrophic for the Dispositional Theory (although one could never become Warranted in holding a belief to that effect, since gaining a Warrant for such a belief would have to proceed through reasoning!). Fortunately this is not so. The principle of Acquisitionary Transmission is not universally invalid. But it is almost universally valid. As a rule of thumb, Acquisitionary Transmission is valid in all cases bar those where it attempts to support some background belief on the basis of some particular empirical fact—Moore's famous proof for the external world being a case in point.

Needless to say, 'rule of thumb' is not a satisfactory means of demarcating valid from invalid instances of Acquisitionary Transmission. We can do better, although it is by no means a straightforward matter exactly how to do so. First, however, it is worth observing what consequences it has for the architecture of one's web of beliefs to lay down such a constraint on Acquisitionary Transmission.

Schematically, one should conceive of one's actual and potential beliefs as a complicated globe-shaped network of logically associated propositions.¹⁹¹ In the lower part of this network are all the Warranted beliefs, while the beliefs that are not yet Warranted are situated on the upper part. Given Closure, one can now work in either of two directions. One can begin from the lower part of the globular network where one's Warranted beliefs are located and work upwards via all logically associated beliefs by *modus ponens* applications of Closure until one (ideally) reaches the top and thus has conferred Warrants on the entire network of beliefs. Or, alternatively, one can begin from the upper part among the beliefs that are not Warranted and via *modus tollens* applications of Closure work downwards until one has deprived oneself of the Warrants for every single belief. So far there is a perfect symmetry. As far as Closure is concerned there is a perfect symmetry here. With Transmission things are slightly different. Transmission allows one to work upwards by Transmitting Warrants from Warranted beliefs to logically associated but not yet Warranted beliefs. There is no negative equivalence to this principle, Distransmission, which destroys Warrants for assumptions whose consequences one lacks Warrants for believing. In this sense, once a given belief is Warranted, it remains so. By inference one can only move upwards, from Warranted beliefs to yet-to-be Warranted beliefs. Naturally, one can by inference gain Warrants that overrule existing Warrants. But that too, is only a constructive

¹⁹¹Why globular and not flat as a classical Quinian web of belief? The third dimension is needed to symbolise the constraints we are laying down on Acquisitionary Transmission. Given such a constraint one cannot, as in a Quinian web of belief, move along any path and in any direction on the web when applying Transmission. Certain beliefs must thus get their Warrant from some other source than Transmission. And even those beliefs that can be Warranted by means of Transmission may only have their Warrant transmitted from a restricted class of beliefs. We can symbolise this by the third dimension: the beliefs constituting the web are ordered such that Warrant only can be transmitted upwards and sideways. Never downwards.

move improving on one's overall epistemic situation. All epistemic methods, including reasoning, are aimed at improving one's overall epistemic situation, either qualitatively, by substituting old Warrants for new and stronger overriding Warrants, or quantitatively by expanding the stock of Warranted beliefs. When reasoning, one starts in one's belief network where one feels one can get a good grip, sufficiently low in the system for there to be recognisably Warranted beliefs floating around, and then moves upwards.

Transmission is an expression of intellectual methodology whereas Closure is an expression of meta-methodology. Closure is not a tool in the epistemological toolbox in the sense that it is not applied in the epistemic project. Rather it is applied in evaluating that project. Closure thus always has an idealisation built into it: once the epistemic project has been *completed*, the result must comply with it. If it did not have this idealisation built into it, it would preclude all intellectual progress. Every time one makes a step forward by reasoning, someone could always insist that it really was a step backwards by a *modus tollens* application of Closure: since the conclusion just drawn was not Warranted prior to the reasoning, neither were the assumptions—thus there was no Warrant in the first place to be transmitted! But this obviously would be an illegitimate move. Transmission states what is knowable by means of inference. Closure states that once the knowable has become known, the overall pattern of one's network of beliefs should comply with Closure.

Bearing in mind these remarks, it emerges that any restriction imposed on the principle of Acquisitionary Transmission is equivalent to defining a certain order of progress in the epistemic project: Transmission states that Warrants can be transported along any path and in any *direction* along a path of logically associated beliefs in one's belief system just as the momentum of a falling dominoes can carry over to (spatially) related dominoes in the row. To state that Warrants cannot *initiate* in this manner from Transmission is merely to claim that this

transportation of Warrants by Transmission has to proceed in a particular order, or direction if you like.

The transportation of Warrants must have a defined initial base. Beliefs belonging to this base must themselves be initially Warranted from some other source than Transmission. Is there such a base? In one sense this question is trivial: is there a starting point in one's epistemic project? There is always both a *de facto* starting point in an epistemic project and a *de facto* order of progression. But this is just to say that one's network of beliefs does not suddenly become Warranted *en bloc* once and for all. It always happens stepwise. However, imposing a constraint on the principle of Acquisitionary Transmission can hardly be considered an attempt to capture a mere *de facto* order in the epistemic project in this trivial sense. Rather it is an attempt to capture a normative order that every epistemic project must comply with and, a *fortiori*, define a base of beliefs that have gained their Warrant independently of, and prior to, transmission of any Warrant.

Imposing a constraint on Acquisitionary Transmission is thus a version of the epistemic policy Michael Williams has termed Epistemological Realism, the view that (i) there exist various *kinds* of beliefs (beliefs about other minds, about the physical world, about experience of the physical world, and so on) such that every belief can be classified as belonging to a particular class exclusively in virtue of its propositional content; and (ii) that these classes are ordered in a fixed epistemological structure such that proper justification of beliefs belonging to one class (e.g. beliefs about the physical world) without exceptions are supported by beliefs belonging to some other class (e.g. beliefs about experience of the physical world).¹⁹² Foundationalism can now be seen to be a classical version of Epistemological Realism. Various categories of beliefs—beliefs about one's one

¹⁹²See Williams 1996, Ch. 3. It is Epistemic Realism, Williams claims, more than metaphysical or truth-theoretical considerations, that is the real source of scepticism.

mental life, mathematical and other *a priori* beliefs and beliefs about experience—are basic and self-warranted and all other categories of beliefs have, one way or another, eventually to be epistemically supportable from either of these classes of fundamental beliefs. But the base of fundamental beliefs suggested by classical foundationalism is not suitable for our purposes. Beliefs about experience might be fundamental in one sense—but not in the sense intended by constraining Acquisitionary Transmission. For surely one can gain a Warrant for the beliefs classical foundationalism considers basic by means of Acquisitionary Transmission.¹⁹³

5.8 Aetiological Foundationalism

The dispositional view on knowledge and justified beliefs we have been developing has led us to reject Acquisitionary Transmission as a universally valid principle. As we have seen, this commits us to some version of Epistemological Realism, the view that our beliefs can be divided into fixed categories from their propositional content alone and, furthermore, that these categories of beliefs fit

¹⁹³As in this case: you look through your collection of photographs from last summer's tracking expedition to the Highlands. You pay attention to a photo depicting a breathtaking view from the summit of some mountain. You do not recall that you experienced this view but come to believe that you did by looking at the photo. You are now Warranted that you had such an experience. Add to the story that the photo in question is not yours at all but has somehow sneaked into your collection. In that case you have never before been Warranted in the belief that you experienced the scenery depicted on the photo. You have thus gained a Warrant for a belief about your experience by means of Acquisitionary Transmission: you are Warranted in believing that you took the photo, transmit that Warrant to the consequent belief that you must have experienced the depicted scenery as you took the photo and thus gain a belief for the first time that you had such an experience.

into an epistemic hierarchy where some categories always support other categories.¹⁹⁴ But it still remains to be specified what these various categories are and, in particular, which category of beliefs has priority in this epistemic structure, that is, which category of beliefs cannot itself be initially Warranted by way of Transmission. To demarcate this category of basic beliefs is just the task advertised earlier of specifying more precisely *which* constraints Acquisitionary Transmission should be subjected to.

According to the dispositional view there is a category of beliefs which are more fundamental than any other beliefs, a category of beliefs, furthermore, the Warrants for which it is impossible to acquire by way of Transmission. These are the aetiological beliefs—beliefs concerned with the origin of the various beliefs in our web of beliefs (hence the phrase ‘Aetiological Foundationalism’). We have already come across a number of such aetiological beliefs. In the Truman case, Truman had a belief regarding the aetiology of his belief that it was raining: that this belief was caused by rain (as opposed to a fake rain-machine aetiology). In the Marquis case, Marquis had a belief regarding the aetiology of his belief concerning a certain behaviour: that it originated from, and to some degree reflected, the mindly state of another person (as opposed to an automaton aetiology). Likewise we have a belief regarding the aetiology of beliefs about our experience generally, *viz.*, that they have a worldly aetiology—that they (partly) are causally determined by the surrounding worldly state of affairs, that is, are products of perception (as opposed to, for example, BIV simulation). We also

¹⁹⁴Could it not be that the dispositional theory, like William’s contextualism, allows for a contextualisation of the epistemic hierarchy between the various categories of beliefs? Only if the evaluation of the subjunctive conditionals featuring in the theory is context sensitive, more specifically: if the World Order is contextually determined. We have argued that World Order is statistically determined and thus remains stable across contexts (or at least stays stable across contexts to the same extent as statistics is stable across contexts).

have a belief regarding the aetiology of beliefs about physical regularities, *viz.*, that we hold them courtesy to a universe governed by uniformity. Finally we have a belief regarding the aetiology of modal beliefs *viz.*, that they ultimately originate from and depict alternative states of affairs that really could have obtained (that is, that our modal beliefs are not illusionary misinterpretations of an in fact deterministic world). For each of these aetiological beliefs one finds a corresponding metaphysical belief: beliefs about the mindly aetiology of behaviour is closely related to a metaphysical belief about the existence of other minds. Beliefs about the worldly aetiology of experience is closely related to a metaphysical belief about the existence of an external world, etc. These metaphysical beliefs are, though, epistemically secondary to the aetiological beliefs, and they can be Warranted by way of Transmission. One could for example exist as the only human being in an environment void of fellow humans (brought up by wolves, monkeys or whatever). Some day one meets, for the first time, another embodied mind. One would, presumably, eventually infer, and thus gain a Warrant, from the behaviour of this creature that there was another mind. But only due to a pre-existing Warrant for the corresponding aetiological belief.¹⁹⁵ If, on the other hand, the metaphysical belief, that there are other minds were the

¹⁹⁵These considerations depend for their plausibility on a particular view on the process of becoming Warranted, more specifically that this involves a two-place relation: between the proposition to be Warranted and a total state of information (web of beliefs). It might be objected that this process rather should be characterised as a three-place relation: between the proposition to be Warranted, a total state of information *and* an appropriate stack of background beliefs. This distinction can, however, be considered as merely terminological, a matter of whether certain background beliefs should be considered part of one's total informational state or constitute a privileged category of beliefs independent of the total state of information. The outcome will be the same: demarcation of a certain sub-set of ones beliefs by defining certain constraints on the principle of Acquisitionary Transmission

epistemologically prior one, why should one ever begin to search for these minds in other behaving bodies—rather than in bicycles, say?

There are local as well as global aetiological beliefs. One such local belief is concerned with the cloudy aetiology of rain beliefs. Another, from the Zebra case, is concerned with the genuine aetiology of beliefs about zebras. Such local aetiological beliefs are more basic than more mundane empirical beliefs, and usually Transmission of Warrants proceeds *from* such beliefs together with Warrants for belief about experience *to* beliefs about empirical states of affairs. Usually—but in some few cases the order is reversed: one might have independent Warrant for the empirical facts (that the creatures in the pen are zebras, say) and one might also have a Warrant for a belief about experience (these creatures appear zebra-like). One then might proceed, via Transmission, to gain a Warrant for the local aetiological belief (these zebra appearances are produced by zebras—and not, for instance, by cleverly disguised mules). The global aetiological beliefs, on the other hand, are prior in a more absolute sense: transmission of Warrants *always* proceeds from global, aetiological beliefs to beliefs about more mundane matters. Hence, Warrants for such beliefs can never be acquired by way of Transmission.

The overall epistemological picture emerging from Aetiological Foundationalism is this: a globular network of beliefs with the global aetiological beliefs at the base. Just above the base is a layer of local aetiological beliefs. By Transmission one can proceed upwards (and sideways) as the epistemological project proceeds. But never downwards.

5.9 Summary

We opened this chapter with a review of the problem cases for the DN Theory. It turned out that some of these dissolved once considered in the light of the principle of Variable Strictness and the principle of Compositionality Strictness as

discussed in Chapter IV. So for instance with the Light case: on the face of it, Flip knows P—that the Light is on next door—according to the Tracking Theory although she does not know this according to common sense intuitions. However, it turned out the subjunctive conditionals in the Tracking Theory are only satisfied if there is a non-accidental relation between execution of Flip's P-investigation—coin-tossing—and P. And, given such a non-accidental relation, Flip's epistemic method—watching the coin—is likened to watching a switch; a method the successful execution of which we usually are happy to associate with knowledge. Other problem cases for the DN Theory persisted despite further scrutiny. We discovered a demarcating feature for these persisting counterexamples, namely that the masking, mimicking and altering mechanisms involved are all closely associated with those worldly states of affairs which the epistemic project is aimed at, whereas those masking, mimicking and altering mechanisms involved in the dissolved counterexamples are all closely associated with the epistemic agent. This reminded us of the dual character of response dispositions discussed in Chapter II: an internal disposition of some agent to respond with a particular belief when confronted with particular facts is always accompanied by an external disposition of those facts to evoke a particular response in the agent. And, as the killer-yellow story brought out, the two dispositions are not always synchronised. According to the CA analysis the latter, the external disposition, is captured by the Tracking conditionals whereas the former, the internal disposition, is captured by the reversed set of conditionals which we termed the Tracker conditionals. We argued that knowledge is execution of an internal dispositional skill rather than manifestation of an external dispositional property. This led us to the formulation of the Dispositional Theory which we stated in terms of Tracker conditionals rather than Tracking conditionals and which provides both an account of knowledge and of Warranted beliefs.

We concluded that with these two improvements in place: an account of subjunctive semantics based on the principle of Variable Strictness and the

principle of Compositionality Strictness together with our reformulation of the theory in terms of the Dispositional Theory, *all* the problem cases have been accounted for. Indeed, our investigations have right from the outset been guided by striking features of the various categories of counterexamples to the DN Theory: in Chapter II we displayed the deficiencies of the DN Theory, its various counterexamples and its counterintuitive rebuttal of Closure. In Chapter III we disclosed the taxonomy of these counterexamples and discovered the affinity with the taxonomy of (alleged) counterexamples to the CA interpretation of dispositional properties, namely the division into instances of masking, mimicking and two sorts of altering. This resemblance in taxonomy between, on the one hand, counterexamples to the CA analysis of dispositions and, on the other hand, the counterexamples to the Tracking analysis of knowledge strongly suggested that knowledge ought to be associated with a dispositional skill whose CA equivalent—the Tracking conditionals—provide a good, although not perfect, approximation. In Chapter IV we then explored the nature of masking, mimicking and altering. The aim here was to develop an improved version of the CA analysis which had the theoretical strength to deal satisfactorily with masking, mimicking and altering generally and, in particular, with instances of masking, mimicking and altering of epistemic dispositionality. In the course of this investigation we discovered that each one of the alleged counterexamples to the CA analysis failed to comply either with the principle of Variable Strictness or with the principle of Compositionality Strictness: in order to satisfy the principle of Variable strictness a non-accidental relation has to be established between the masking, mimicking or altering mechanism and the entity with the dispositional property. But the only manner that can be accomplished is by attaching the mechanism in question to the object with the dispositional property and hence construct a composed object the overall dispositional properties of which are distinct from those of the original object (cf. the principle of Compositionality Strictness). However, when we applied these results in the present chapter to the corresponding CA analysis of

epistemic dispositional, the DN theory, it turned out that only those counterexamples dissolved where the masking, mimicking and altering mechanism was non-accidentally associated with the epistemic agent and her execution of some epistemic method. Counterexamples where these mechanisms were non-accidentally connected with P-facts, on the other hand, persisted. This fact reminded us that there are two, and not just one, dispositions involved in the epistemic project: the internal dispositional skill of the agent to form true beliefs when confronted with P-matters and the external disposition of P-matters to evoke particular beliefs in the agent. So, here again, it was a specific feature of the remaining counterexamples that initially motivated the idea that the epistemic competence has hitherto fallaciously been associated with the external disposition of the world rather than with the internal disposition of the epistemic agent. Indeed, careful reading of the CA analysis reveals that the internal disposition of the agent should be associated with the Tracker conditionals rather than with the Tracking conditionals.

But it is important to emphasise that although it is taxonomic and other features of the counterexamples to the DN Theory which guided and initially motivated the principles underlying the Dispositional Theory and, in particular the two issues on which it diverges from the DN Theory (reversal of the subjunctive conditionals and a semantics for those conditionals guided by the principle of Variable Strictness and the principle of Compositionality Strictness) nothing essential hinges on the fact that the Dispositional Theory really can provide the correct verdict for all present and future instances of knowledge acquisition. The present work has not been guided by an urge to develop a technical theory for which no counterexamples can be found. The purpose has been to learn from these counterexamples, to gain new insights about man's place in the world, in particular the nature of our epistemic link to various subject matters and the principal limit governing such a link. Although the initial motivation for the various components in the Dispositional Theory has been provided by a thorough

analysis of counterexamples to the DN Theory, it has turned out, subsequently, that the suggested changes have an independent rationale. So for instance with the suggestion that the Tracker conditionals rather than the Tracking conditionals provide a suitable measure for the agent's epistemic skill. According to CA, the Tracker conditionals characterise an internal disposition whereas the Tracking conditionals characterise an external disposition. And although our epistemic link to the world involves both internal and external properties, it is the internal agency that is most relevant when evaluating the epistemic skill of an epistemic agent. We can thus say that whatever the theoretical merits of epistemic dispositional theories generally, *ceteris paribus* it will always be preferable to have the focus moved from external to internal dispositionality. Likewise for our results regarding subjunctive semantics: Whatever their merit of the CA analysis of epistemic and other dispositions, they offer clarification of some potentially confusing issues that go right to the heart of subjunctive semantics and are thus valuable independently of their contribution to the Dispositional Theory.

With these points in mind we ought also to consider the last part of the present chapter where we proceeded to a dispositional diagnosis of some of the classic problem cases for the DN theory and the sceptical challenge vis-à-vis the Dispositional Theory. I postulated that it is a great virtue of the Dispositional Theory that it provides a sane response to the sceptical Trilemma, namely by granting us knowledge that some particular sceptical scenario has not materialised. This as opposed to other recent works on epistemology where the line is to reject either the principle of Closure or knowledge of mundane empirical facts—be it *en bloc* or contextually selected portions thereof. Again, even if this postulate does not convince the reader, that is, if one doubts our findings offer an efficient rebuttal of scepticism, it is important to emphasise that it still uncovers some new anti-sceptical strategies to be explored further—some strategies that are foreclosed in, for instance, the DN theory. It is fairly certain that one does not know one is not a BIV according to the DN Theory since the first Tracking conditional is not

satisfied. According to the Dispositional Theory, on the other hand, there is still latitude to argue that the Tracker conditionals as a matter of fact are satisfied for such a knowledge claim. Hence, I can leave it as an open question whether or not one does know such things according to the Dispositional Theory and still claim some progress over the DN Theory according to which it is a decisive fact that we do not know these things.

We closed the chapter by a brief outline of an epistemological position we phrased ‘aetiological foundationalism’. We distinguished the principle of Closure, which I wish to claim holds according to the Dispositional Theory, from the principle of Acquisitionary Transmission which I argued fails according to the Dispositional Theory. That is, I argued that some beliefs are epistemically prior in the epistemological project in the sense that they cannot initially be Warranted by way of Transmission but rather serve as a foundation on which the epistemic project must be based. But I argued that the beliefs that constitute this foundation are not, as often assumed, metaphysical in nature. Against this I claimed that they concern the aetiology of the cognitive processes that take place when we form our various beliefs about the world.

Bibliography

Annis, D. (1978) 'A Contextualist Theory of Epistemic Justification'

American Philosophical Quarterly (vol.28)

Armstrong, D. (1973) *Belief, Truth and Knowledge*

Cambridge: CUP

Barker, S. (1957) *Induction and Hypothesis*

New York: Cornell

Bird, A. (1998) 'Dispositions and Antidotes' (vol.48)

The Philosophical Quarterly

Blackburn, S. (1969) 'Goodman's Paradox', in N. Rescher (ed.)

Studies in Philosophy of Science

— (1973) *Reason and Prediction*

Cambridge: CUP

— (1993a) *Essays in Quasi-Realism*

Oxford: OUP

— (1993b) 'Knowledge, Truth and Reliability', in Blackburn (1993a)

— (1993c) 'Circles, Finks, Smells and Biconditionals'

Philosophical Perspectives (vol. 7)

Boghossian, P. and Velleman, J. (1989) 'Colour as a Secondary Quality'

Mind (vol.98)

Brueckner, A. (1984) 'Why Nozick is a Sceptic'

Mind (vol. 93)

— (1985) 'Losing Track of the Sceptic'

Analysis (vol.45)

- (1991) 'Unfair to Nozick'
Analysis (vol.51)
- (1997) 'Closure and Context'
Ratio
- Carnap, R. (1946) 'On the Application of Inductive Logic'
Philosophy and Phenomenological Research (vol.6)
- Coady, C.(1992) *Testimony*
 Oxford: Clarendon Press
- Craig, E. (1989) 'Nozick and the Sceptic: The Thumbnail Version'
Analysis (vol.49)
- (1990) *Knowledge and the State of Nature*
 Oxford: Clarendon Press
- (1993) 'Understanding Scepticism', in Haldane and Wright eds.
Realism, Representation and Projection, Oxford: OUP
- Dancy, J. (1984) 'On the Tracks of the Sceptic'
Analysis (vol.44)
- (1985) *Introduction to Contemporary Epistemology*
 Oxford: Blackwell
- Davies, D. (1997) 'Why One Shouldn't Make an Example of a Brain in a Vat'
Analysis (vol.57)
- Davies, M. (1999) 'Externalism, Architecturalism and Epistemic Warrant', in
On Knowing One's Mind, C. Wright et. al. (eds.)
- DeRose, K. (1995) 'Solving the Skeptical Problem'
Philosophical Review
- (1996) 'Knowledge, Assertions and Lotteries'
Australasian Journal of Philosophy (vol.84)
- Dretske, F. (1970) 'Epistemic Operators'
Journal of Philosophy (vol.67)

- (1971) 'Conclusive Reasons'
Australasian Journal of Philosophy (vol.49)
- (1981) *Knowledge and the Flow of Information*
 Oxford: Blackwell
- (1997) 'Pyrrhonian Reflections on Knowledge and Justification'
Philosophy and Phenomenological Research (vol.57)
- Dummett, M. (1993) *The Seas of Language*
 Oxford: Clarendon Press
- (1993a) 'Testimony and Memory' in
 Dummett (1993)
- (1997) 'Gifford Lectures'
 Unpublished manuscript
- Forbes, G. (1984) 'Nozick on Scepticism'
The Philosophical Quarterly (vol.34)
- (1995) 'Realism and Skepticism: Brains in a Vat Revisited'
Journal of Philosophy (vol.102)
- Garrett, B. (1983) 'Nozick on Knowledge'
Analysis (vol.43)
- Gettier, L. (1963) 'Is Justified True Belief Always Knowledge?'
Analysis (vol.23)
- Goldman Alan. (1986) 'Nozick on Knowledge: Finding the Right Connection', in
 Lupper-Foy (1986)
- Goldman, Alvin. (1967) 'A Causal Theory of Knowing'
Journal of Philosophy (vol.64)
- (1973) 'Discrimination and Perceptual Knowledge'
Journal of Philosophy (vol.70)
- Goodman, N. (1946) 'A Query on Confirmation'
Journal of Philosophy (vol.53)

- (1947) ‘On Infirmities of Confirmation-theory’
Philosophy and Phenomenological Research (vol.7)
- (1954) *Fact, Fiction and Forecast*
Sussex: Harvester Press
- (1963) ‘Faulty Formalization’
Journal of Philosophy (vol.60)
- (1966a) ‘Foreword’
Journal of Philosophy (vol.63)
- (1966b) ‘Comments’
Journal of Philosophy (vol.63)
- Gundersen, L. (1997) Review of McCall: *A Model of the Universe*
The Philosophical Quarterly (Vol.47)
- (2000a) ‘Bird on Dispositions and Antidotes’
The Philosophical Quarterly (Vol.50)
- (2000b) ‘Goodman’s Gruesome Modal Fallacy’
Facts, Things & Events, Faye et al. (eds.), Puznan Studies
- Hacking, I. (1965) *The Logic of Statistical Inference*
New York: Cambridge
- (1993) ‘On Kripke and Goodman’s Uses of ‘Grue’’
Philosophy
- Hale, B. (1988) ‘Epistemic Universalizability’
Analysis (vol.48)
- Harman, G. (1968) ‘Knowledge, Inference and Explanation’
American Philosophical Quarterly
- Hetherington, S. (1998) ‘Actually Knowing’
The Philosophical Quarterly (vol.48)
- Heidegger, M. (1962) *Being and Time* (tran.)
Oxford (first published 1927)

- Hume, D. (1967) *An Inquiry Concerning Human Understanding* (C. Hendel ed.)
New York: Bobbs-Merrill (first published 1739)
- Jackson, F. (1975) 'Grue'
Journal of Philosophy (vol.72)
- (1996) 'The Primary Quality View of Color'
Philosophical Perspectives (vol.10)
- Johnston, M. (1989) 'Dispositional Theories of Value'
Proceedings of the Aristotelian Society
- (1991) 'Explanation, Response-Dependence, and Judgement-Dependence'
Response-Dependent Concepts, Menzies (ed.) ANU, Working Papers in
Philosophy
- (1992a) 'How to Speak of the Colors'
Philosophical Studies (vol.46)
- (1992b) 'Remarks on Response-Dependence'
Unpublished manuscript
- (1993) 'Objectivity Refigured: Pragmatism without Verificationism'
Reality, Representation and Projection (Haldane and Wright eds.)
- (1999) 'Are Manifest Qualities Response-Dependent?'
The Monist
- Kant, I. (1933) *Critique of Pure Reason* (Kemp Smith tran.)
London: MacMillan Press (first published 1787)
- Kyberg, H. (1961) *Probability and the Logic of Rational Belief*
London: Harvard University Press
- Lewis, D. (1979) 'Counterfactuals, Dependence and Time's Arrow'
Nous (vol.13)
- (1989) 'Dispositional Theories of Value'
Proceedings of the Aristotelian Society
- (1996) 'Elusive Knowledge'
Australasian Journal of Philosophy (vol.74)

- (1997a) 'Finkish Dispositions'
The Philosophical Quarterly (vol.47)
- (1997b) 'Naming the Colours'
Australasian Journal of Philosophy (vol.75)
- Luper-Foy, S. (1984) 'The Epistemic Predicament: Knowledge, Nozickian Tracking and Scepticism'
Australasian Journal of Philosophy (vol.62)
- (1987) (ed.) *The Possibility of Knowledge*
 Rowman & Littlefield
- McCall, S. (1994) *A Model of the Universe*.
 Oxford: Clarendon Press
- McDowell, J.(1882) 'Criteria, Defeasibility and Knowledge'
Proceedings of the British Academy
- (1994) *Mind and World*
 London: Harvard University Press
- (1995) 'Knowledge and the Internal'
Philosophy and Phenomenological Research (vol.55)
- McGinn, C. (1984) 'The Concept of Knowledge'
Midwest Studies in Philosophy
- McGinn, M. (1989) *Sense and Certainty*
 Oxford: Blackwell
- Martin, C. (1994) 'Dispositions and Conditionals'
Philosophical Quarterly (vol.44)
- Menzies, P. & Pettit, P. (1993) 'Found: the Missing Explanation'
Analysis (vol.53)
- Miller, A. (1992) 'Some Responses to the Missing Explanation Argument'
 Mimeo, University of Michigan
- (1995) 'Objectivity Disfigured: Mark Johnston's Missing-Explanation Argument' *Philosophy and Phenomenological Research*, (Vol. 55)

- (1997) 'More Responses to the Missing-Explanation Argument'
Philosophia (Vol.25)
- Nozick, R. (1981) *Philosophical Explanations*
Oxford: Clarendon Press
- Pettit, P. (1991) 'Realism and Response Dependence'
Mind (vol.100)
- Plantinga, A. (1993a) *Warrant: The Current Debate*
Oxford: OUP
- (1993b) *Warrant and Proper Functioning*
Oxford: OUP
- Prior, E. (1985) *Dispositions*
Aberdeen: Aberdeen University Press
- Prior, E., Pargetter, R., Jackson, F. (1982) 'Three Thesis About Dispositions'
American Philosophical Quarterly
- Putnam, H. (1982) 'Brains in a Vat', in his
Reason, Truth and History, Cambridge: CUP
- Quine, O. (1969) 'Epistemology Naturalized', in his
Ontological Relativity and Other Essays (New York: Columbia UP)
- Rorty, R. (1979) *Philosophy and the Mirror of Nature*
Princeton: PUP
- Russell, B. (1948) *Human Knowledge: Its scope and Limits*
London: Allen and Unwin Ltd
- Schiffer, S. (1996) 'Contextualist Solutions to Scepticism'
Proceedings of the Aristotelian Society
- Sen, P. (1997) 'S. K. Bose Memorial Lectures'
Unpublished manuscript
- Shope, R. (1978) 'The Conditional Fallacy in Contemporary Philosophy'
Journal of Philosophy (vol.75)

- (1984) ‘Cognitive Abilities, Conditionals and Knowledge’
Journal of Philosophy (vol.81)
- Stalnaker, R. (1968) ‘A Theory of Conditionals’ in Rescher (ed.)
Studies in Logical Theory, Oxford: Blackwell
- Strawson, P. (1985) *Skepticism & Naturalism - Some Varieties*
 London: Methuen
- Stroud, B. (1984) *The Significance of Philosophical Scepticism*
 Oxford: Clarendon Press
- Unger, P. (1975) *Ignorance: A Case for Scepticism*
 New York: OUP
- Vogel, J. (1987) ‘Tracking, Closure and Inductive Knowledge’, in
 Lupper-Foy ed. (1987)
- Williams, M. (1977) *Groundless Beliefs*
 Oxford
- (1996) *Unnatural Doubts*
 Princeton University Press
- (1993) ‘Realism and Scepticism’, in Haldane and Wright eds.
Realism, Representation and Projection, New York: OUP
- Wittgenstein, L. (1979) *On Certainty* (G. Anscombe and G. von Wright eds.)
 Oxford: Blackwell
- Wright, C. (1983) ‘Keeping Track of Nozick’
Analysis (vol.43)
- (1986) ‘Facts and Certainty’
Proceedings of the British Academy
- (1991a) ‘Order of Determination, Response-Dependence, and the Euthyphro Contrast’
 Unpublished manuscript
- (1991b) ‘Scepticism and Dreaming: Imploding the Demon’
Mind (vol.100)

- (1992a) *Truth & Objectivity*
London: Harvard University Press
- (1992b) ‘On Putnam’s Proof that We Are not Brains in a Vat’
Proceedings of the Aristotelian Society
- (1993a) *Realism, Meaning & Truth*
Oxford: Blackwell
- (1993b) ‘Realism: The Contemporary Debate—W(h)ither Now?’
Reality, Representation and Projection (Haldane and Wright eds.)
- (1993c) Lecture notes for ‘Advanced Epistemology’
Unpublished manuscript
- (1997) ‘On Acquisition of Warrant by Inference’
Unpublished manuscript
- (1998a) Lecture notes on Response- and Judgement-dependence
Unpublished manuscript
- (1998b) ‘Cogency and Question-Begging: Some Reflections on McKinsey’s
Paradox and Putnam’s Proof’, Unpublished manuscript
- (1998c) Lecture notes for ‘Columbia Seminar’
Unpublished manuscript
- Zabludowski, A. (1974) ‘Concerning a Fiction About How Facts Are Forecast’
Journal of Philosophy (vol.81)