The Goodman-Kripke Paradox

Robert Kowalenko

King's College London

PhD
Philosophy

2003

Abstract

The Kripke/Wittgenstein paradox and Goodman's riddle of induction can be construed as problems of multiple redescription, where the relevant sceptical challenge is to provide factual grounds justifying the description we favour. A choice of description or predicate, in turn, is tantamount to the choice of a curve over a set of data, a choice apparently governed by implicitly operating constraints on the relevant space of possibilities. Armed with this analysis of the two paradoxes, several realist solutions of Kripke's paradox are examined that appeal to dispositions or other non-occurrent properties. It is found that all neglect crucial epistemological issues: the entities typically appealed to are not observational and must be inferred on the basis of observed entities or events; yet, the relevant sceptical challenge concerns precisely the factual basis on which this inference is made and the constraints operating on it. All disposition ascriptions, the thesis goes on to argue, contain elements of idealization. To ward off the danger of vacuity resulting from the fact that any disposition ascription is true under just the right ideal conditions, dispositional theories need to specify limits on legitimate forms of idealization. This is best done by construing disposition ascriptions as forms of (implicit) curve-fitting, I argue, where the "data" is not necessarily numeric, and the "curve" fitted not necessarily graphic. This brings us full circle: Goodman's and Kripke's problems are problems concerning curve-fitting, and the solutions for it appeal to entities the postulation of which is the result of curve-fitting. The way to break the circle must come from a methodology governing the xidealizations, or inferences to the best idealization, that are a part of curve-fitting. The thesis closes with an argument for why natural science cannot be expected to be of much help in this domain, given the ubiquity of idealization.

Table of Contents

Introduction

In his foreword to the fourth edition of Nelson Goodman's *Fact, Fiction and Forecast*[1] Hilary Putnam points to a strong resemblance between Goodman's treatment of induction and the later Wittgenstein's considerations on rule-following. The resemblance obtains on a particular interpretation of Wittgenstein, as put forward in Saul Kripke's well-known commentary *Wittgenstein on Rules and Private Language.*[2] There, Kripke famously suggests that Wittgenstein ought to be considered the father of a new form of philosophical scepticism founded on a paradox about rule-following and meaning. Wittgenstein's alleged scepticism 'should be obvious to any reader of Goodman,' Kripke points out, for Goodman's strategy in deploying the "new riddle of induction" is strikingly close to Wittgenstein's sceptical arguments (Kripke 1982, pp. 20, 58). Although he is not the first to make observations of this kind, Kripke's exegesis proved particularly influential, generating, in the words of one commentator, 'excitement unparalleled since the heyday of Wittgenstein scholarship in the early 1960s.'[3] Kripke's interpretation prompted a vast amount of new work on the (by that time) well-worn subject of rule-following, and even philosophers not normally concerned with Wittgenstein's views took interest in the new form of meaning scepticism put forward. However, although some authors refer to Wittgenstein's alleged sceptical stance in one breath with Goodman's treatment of induction,[4] Kripke remains to date the only distinguished philosopher in the field to une-

---

[1] First published as Goodman, N. (1954). *Fact, Fiction, and Forecast*, London, Athlone Press. All page references will be to the fourth edition, Goodman, N. (1983). *Fact, Fiction, and Forecast*, Cambridge, Harvard University Press. Putnam, incidentally, credits Catherine Elgin with having suggested the resemblance to him.

[2] Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language*, Oxford, Blackwell.

[3] Boghossian, P. A. (1989). "The Rule-Following Considerations" *Mind* **98**: 507-549, at p. 507. Earlier remarks on the affinity between Goodman and Wittgenstein can be found in Blackburn, S. (1969). "Goodman's Paradox" *Studies in the Philosophy of Science*. N. Rescher, Oxford, Blackwell, **3:** 128-42; Hacking, I. (1975). *Why Does Language Matter to Philosophy?*, Cambridge University Press, Cambridge, p. 69.

[4] E.g. Millikan, R. G. (1990). "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox" *Philosophical Review* **99**(3): 323-53, p. 334; Stroud, B. (1990). "Meaning, Understanding and Translation" *Canadian Journal of Philosophy* **16**: 343-361; Mellor, D. H. (1995). *The Facts of Causation*, New York, Routledge, p. 32; Martin, R. M. (1990). "It's Not that Easy Being Grue" *Philosophical Quarterly*: 40(160) 299-315. Sainsbury, R. M. (1995). *Paradoxes*, New York, Cambridge University Press, p. 93;. Mulhall, S. (1989). "No Smoke without Fire: The Meaning of Grue" *Philosophical Quarterly* **39**: 166-189; Stegmüller, W. (1989). *Hauptströmungen der Gegenwartsphilosophie (Vol. 4)*,

quivocally link Goodman's famous 'new riddle of induction' to Wittgenstein's purported scepticism about meaning. Kripke writes: 'Although Goodman concentrates on the problem about induction and largely ignores the problem about meaning, his discussions are occasionally suggestive for Wittgenstein's problem as well. In fact, I personally suspect that *serious consideration of Goodman's problem, as he formulates it, may prove impossible without consideration of Wittgenstein's*' (Kripke 1982, p. 59; my emphasis).

The significant connection between two important philosophical problems adumbrated in this quote shall serve as a starting point of the present thesis. A close relationship, if substantiated, between the new meaning scepticism and sceptical considerations relative to Goodman's riddle of induction promises interesting and potentially wide-ranging philosophical consequences. Before we take a look at these, however, a preliminary remark is in order. As was to be expected, Kripke's elegant and original, yet perhaps slightly cavalier interpretation of Wittgenstein has generated considerable controversy among commentators, even indignation in some quarters, for a perceived lack of exegetical effort and precision.[5] We shall throughout this thesis steer clear of that particular debate and—in the image of much previous work on "Kripkenstein"—not address the question whether or to what extent Kripke's Wittgenstein is the actual Wittgenstein. For the particular brand of meaning scepticism Kripke has put on the table, be it an act of faithful exposition or original philosophizing vaguely inspired by Wittgensteinian themes, deserves consideration on its own merits. A prior analysis of Kripke's exegetical achievements and/or failures would in fact be superfluous, given our intention to contrast specifically Kripke-style scepti-

Stuttgart, Alfred Kröner Verlag, pp. 30-34. Blackburn, S. (1984b). *Spreading the Word: Groundings in the Philosophy of Language*, Oxford, Clarendon Press devotes a whole section to 'Wittgenstein and Goodman: „bent" predicates', but offers little in the way of a closer examination of the relationship between the two.

[5] Cf. Baker, G. P. and P. M. S. Hacker (1984a). "On Misunderstanding Wittgenstein: Kripke's Private Language Argument" *Synthese* **58**: 407-50 and Baker, G. P. and P. M. S. Hacker (1984b). *Scepticism, Rules and Language*, Oxford, Blackwell. Baker and Hacker ironically note that Kripke must have used Wittgenstein's writings as a Rorschach spot, whereas Savigny, E. V. (1988). *Wittgensteins "Philosophische Untersuchungen": Ein Kommentar für Leser (Band 1)*, Frankfurt, Klostermann, scathingly accuses him of providing a poor example of *irresponsible scholarship* to new generations of graduate students. Other commentators are more measured but equally firm in the rejection of Kripke's take on Wittgenstein. See e.g. Blackburn, S. (1984a). "The Individual Strikes Back" *Synthese* **58**: 281-302; McDowell, J. (1984). "Wittgenstein on Following a Rule" *Synthèse* **Vol. 58**: 326–363; Shanker, *Wittgenstein and the Turning-Point in the Philosophy of Mathematics*, London: Croom Helm, 1987, pp. 13-25. For rare dissenting voices—authors who believe that Kripke's Scepticism is also Wittgenstein's—see Wilson, G. M. (1998). "Semantic Realism and Kripke's Wittgenstein" *Philosophy and Phenomenological Research* **58**(1): 99-122; also Allen, B. (1989). "Gruesome Arithmetic: Kripke's Sceptic Replies" *Dialogue* **28**(2): 257-264

cism with the new riddle of induction. Thus, although this thesis will be in its first part concerned with a quintessentially Wittgensteinian theme, rule-following and meaning, Wittgenstein himself shall barely enter the stage, only "Wittgenstein as he struck Kripke."

What does Kripke mean, then, with the remark that 'serious consideration of Goodman's problem may prove impossible without consideration of Wittgenstein's'? The very idea may strike one as puzzling, even prior to delving into the details of the respective arguments. If philosophical problem A is impossible to solve without taking into account philosophical problem B, then, one might argue, B is in some sense part of A. This, however, seems unlikely in our case: Goodman's well-known discussion of the 'new riddle of induction' shuns every reference to *meaning*, as Kripke is the first to acknowledge. Goodman never speaks of, say, the meaning of a particular inductive hypothesis, or of the meaning of any sub-sentential parts of it, because he believes that the notion of meaning is, quite generally, philosophically disreputable and should be avoided.[6] Rather, Goodman's new riddle underscores the existence of a peculiar indeterminacy in the relation between certain types of proposition describing our evidence, and the general proposition or law-like statement they are usually thought to "fall under" and confirm.[7] The import of his riddle seems to be that we lack a satisfactory explanation for why we select certain predicates rather than others for use in law-like generalizations about the world, and that we have no satisfactory account of how we *should* select them in order to obtain reliable inferences and predictions. Kripke's paradox, on the other hand, suggests (in cartoon form) that neither I nor even an omniscient God can know just which object I am referring to with the expressions I use, simply because there are no facts of the matter uniquely determining the reference of both linguistic and mental items. It thus apparently entails both scepticism about self-knowledge as well as a certain anti-realism about meaning.

These are rather different-looking issues in different areas of philosophy, and it is quite legitimate to question the plausibility of any suggestion that there could be a substantial link between them—one might even ask why it would so much as *matter* if there were one.

[6] For Goodman's views on the issue of meaning, see Goodman, N. (1949). "On Likeness of Meaning" *Analysis* **Vol. 10**: 1–7 and Goodman, N. (1953). "On Some Differences about Meaning" *Analysis* **Vol. 13**: 90–96.

[7] These brief allusions to the two paradoxes are of course mere advertisements for coming attractions. A full description is given in Sections 1.1 and 1.2, respectively.

The reason it would matter can be stated as follows. Kripke's paradox expresses, in the words of Crispin Wright, 'one of the late twentieth century's most characteristic philosophical preoccupations,'[8] namely scepticism about semantic notions. Wright holds that Kripke's interpretation of the second Wittgenstein is one of the three principal contemporary statements of that position, along with W.V. O. Quine's indeterminacy-of-translation thesis and Hilary Putnam's 'model-theoretic' criticism of realism.[9] A great number of philosophers generally opposed to meaning scepticism in any of its forms have sensed the importance of the new challenge, and volunteered objections to Kripke's argument.[10] Philosophers inclined towards some forms of naturalism or realism in particular tend to straightforwardly dismiss it. Interestingly, however, many of the realist or naturalist criticisms share an important feature—a reliance (in ways more or less explicit) on the assumption that Goodman's paradox[11] has been defused or *can* be defused. This assumption does not at first glance seem particularly worrying, for there is nothing wrong, in philosophy, with building on results established or to be established in an adjacent field of philosophy. There is nothing wrong, in any sort of enquiry, with adopting the strategy of divide and conquer and to assume one problem to be solvable in order to give oneself the means to tackle another. The strategy's correct application demands, however, that the dividing be done at the "joints" of the subject matter under scrutiny. There must, in other words, be something to divide in the first place. The first chapter of the present work shall argue that Kripke's and Goodman's paradox are *essentially con-*

[8] Cf. Wright, C. (1997). "The Indeterminacy of Translation" *A Companion to the Philosophy of Language*. C. Wright, Oxford, Blackwell, p. 397; Puhl, K., (Ed. (1991). *Meaning Scepticism*, Berlin, de Gruyter, pp. 1-2.

[9] Cf. Quine, W. V. O. (1960). *Word and Object*, Cambridge, MIT Press and Putnam, H. (1980). "Models and Reality" *Journal of Symbolic Logic* **45**: 464–482.

[10] Some of these objections turn on Kripke's apparent exegetical imprecision and mount an immanent and purportedly Wittgensteinian critique of the sceptical argument. Most, however, attempt to demonstrate the invalidity of the paradox using external, "non-Wittgensteinian" arguments. For explicitly Wittgensteinian objections, see e.g. Baker and Hacker *Scepticism, Rules and Language* and the collection of articles in Puhl, (Ed. *Meaning Scepticism*. For objections in a more loosely "Wittgensteinian spirit," see e.g. Horwich, P. (1990). "Wittgenstein and Kripke on the Nature of Meaning" *Mind and Language* **5**(2): 105-121, and Horwich, P. (1995). "Meaning, Use and Truth" *Mind* **104**(414): 355-368. For external objections, see e.g. Fodor, J. A. (1990). "A Theory of Content II" *A Theory of Content and other Essays*, Cambridge, Masachusetts, MIT Press; Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*, Ny, Praeger; Millikan "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox", pp. 223-43. Given the volume of the relevant literature, this list is selective.

[11] I follow common practice and refer to what Goodman called 'the new riddle of induction' as 'Goodman's paradox', though some commentators suggest we need not, and should not, think of it as a paradox proper (cf. Shoemaker, S. (1975). "On Projecting the Unprojectible" *Philosophical Review* **84**: 178-219.)

nected, and that therefore we better be circumspect, at least in some types of philosophical context, of divorcing them.

Any theory purporting to show that a given paradox cannot be solved independently from another needs to demonstrate that the philosophical difficulties giving rise to the paradoxes are either (a) in some important sense identical, (b) a direct consequence of each other, or (c) each a direct consequence of one and the same further philosophical problem. It does not, in my view, need to show that the two paradoxes proceed from logically or otherwise equivalent premises. Clearly, however, general considerations simply pointing out a certain 'family resemblance' between the respective arguments do not suffice to establish the desired conclusion—nor does Kripke's rather elliptical remark quoted above. In the first part of this thesis I shall thus attempt to give more substance to Kripke's assertion, in order to show that his intuition (as well as that of some others) concerning the two paradoxes is correct. The two paradoxes are indeed closely related, I argue, because they can both be meaningfully be characterized as riddles about redescription, even curve-fitting. Both Kripke and Goodman make us wrestle with the question whether there are, or should be, any objective constraints on how we choose to describe a series of events, or fit a curve over a series of points. Can we escape the relativism which looms if we allow that these sorts of acts are determined by the way we, subjectively, perceive simplicity and similarity? I do not argue directly in favour or against taking seriously either of the two paradoxes, for I am interested in their relationship, and any conclusion concerning the nature of the connection between the two paradoxes does not prejudice the independent question of their validity. Yet it is clear that if Kripke's and Goodman's problems are in a certain respect two sides of the same philosophical coin, this will deeply affect the path of thought any prospective candidate for their solution must follow.

The second part of this thesis will go on to consider in detail some of the realist/naturalistic responses to Kripke's paradox we have mentioned above. Their central idea throughout is that Kripke's scepticism about meaning is to be answered by appealing to a speaker's *dispositional* properties of a speaker, rather than his occurrent ones, or to some sort of *conditional* facts about her, rather than actual ones. Of course, appeals to dispositions and other non-occurrent states have been a mainstay of philosophical studies of the mind since, at least, Gilbert Ryle's *The Concept of*

Mind.[12] But our contemporary concept of dispositionality has considerably evolved away from the empiricist notion of 'disposition' as it was available to early and mid-20th century behaviourists, and it has done so in rather close synchronisation with other changes in our thinking about the mind and the world. Thus, although dispositions, dispositional states, and closely related concepts such as 'competence', remain today as important to cognitive psychologists, linguists, and naturalistic philosophers as they once were to their behaviourist predecessors—most theoretical accounts of dispositionality now mirror the so-called "mentalistic turn" (away from behaviourism) which took place in these disciplines by effecting what might be described as a corresponding "realist turn" (away from empiricism about dispositions). In the wake of this development, dispositions have re-emerged as possible candidates for the fundamental building-blocks of ontology, after particulars, universals, properties, tropes, events, and states of affairs.

The various candidate solutions to Kripke's paradox, although all explicitly *realist* or *naturalist* and very ostensibly dressed in an ontological attire, are examined from an epistemological point of view here. This is as it should be—for defeating Scepticism is the maxim. Kripke's paradox can be used as a sort of litmus test for the epistemological soundness of the various new realisms about dispositions that are currently on offer; the results are mixed. Kripke's Sceptic is defeated, we find, only at a heavy philosophical price—important assumptions that need to be independently argued for. These assumptions involve the nature of our epistemic access to the dispositional properties or conditional facts that are postulated to achieve the desired purpose of uniquely fixing meaning. Due to the very nature of the entities postulated, this access cannot be direct, or observational. All realists within the purview of this study thus make assumptions, some explicit but most implicit, about what needs to be in place in order to be able to *infer*, in an epistemologically water-tight way, the relevant dispositional entities from what can directly be observed. One of our realists, for example, acknowledges that her refutation of the Sceptic turns on the assumption that there is '...a metaphysical distinction between natural properties and kinds and artificially synthezised grue-like properties and kinds or, what is perhaps the same, depends upon there being a difference between natural law and mere *de facto* regular-

[12] Ryle, G. (1949). *The Concept of Mind*, London, Hutchinson's University Library. Arguably, however, the dispositionalist approach to the mind began with Aristotle (see e.g. Aristotle (1974). *Categories; and, De interpretatione. Translated [from the Greek] with notes by J.L. Ackrill*, Oxford, Clarendon Press, Book I, 8).

ity.'[13] Well, if that is what our victory against the Kripkean Sceptic hinges on, then, given our contention that Kripke's and Goodman's paradox are but two sides of the same coin, we are not doing very well against Scepticism! I propose a synoptic view of all realist solutions is proposed, and argue that because they all trade on claims whose truth-conditions are far from observational, any decision to ascribe the sort of dispositional state or other non-occurrent property that is purported to defeat the Sceptic's hypothesis, are based on the same sort of *theoretical inference* from the observed to the unobserved. Of course, nothing else was to be expected: the "realist turn" in the theory of dispositions followed, as we have noted above, closely on the heels of the mentalist turn in the philosophy of mind, whose central feature was precisely the jettisoning of all behaviourist scruples with regard to the legitimacy of inferences from the observable to the unobservable.

A new twist is introduced into our story by those realists about dispositions, such as Jerry Fodor, who are perfectly aware of this and acknowledge that viable disposition ascriptions to objects can only be made via scientific laws that are, on the one hand, defined over *ideal* conditions, and on the other, true only *ceteris paribus*. The notion of 'idealization' and 'ceteris paribus clause' thus take centre stage in the final parts of this study, which has taken a turn from the philosophy of language to metaphysics, to the philosophy of science. I describe how many *ceteris paribus* claims contain implicit idealizations, and how many idealizations consist in ignoring potential troublemakers, in other words, in assuming that everything remains "*ceteris paribus*." To unify the phenomena, I propose a hypothesis according to which disposition-ascriptions, idealized laws, and ceteris-paribus laws all have the same conditional form with a description of non-actual conditions in the antecedent, and that they are obtained through the same sort of inference. This inference seems to involve simplicity and similarity considerations, and we are back full circle to the problems that have occupied us at the beginning. The final conclusions of this thesis will be more speculative than the present author would have like them to be. But, the reader will appreciate, Sceptical challenges are among the most recalcitrant challenges of philosophy. We must be content if small progress is made against the Sceptic; the author's hope is that this study will contribute to that effort.

Further talking about these issues without introducing our two paradoxes would be like talking about 'Hamlet' without the Prince, however. Enter the Prince.

---

[13] The naturalistic philosopher of mind, Ruth Garrett Millikan in Millikan "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox", p. 334.

1. Two Paradoxes

This Chapter begins with a characterization of what I take to be the essential features of the two paradoxes we shall be concerned with. Inevitably, this involves mention of arguments which ‘have been discussed so often that the utility of yet another exposition is certainly open to question’ (Kripke). Instead of once more belabouring all-to-familiar ground, the discussion to follow shall rather attempt to paint what Wittgenstein would call a *übersichtliche Darstellung* (conspicuous representation), aimed at bringing out the common features of Goodman’s and Kripke’s paradox. For expository reasons, I reverse chronological order and begin with Kripke.

Section 1.1 presents a formalized version of the paradox as put foward by Kripke 1982. Section 1.1.1 follows Boghossian 1989 in arguing that Kripke’s Sceptic effectively rejects all relevant candidates for being a meaning-constitutive fact because they fail one or both of two tests: meaning is (a), *normative*, and (b) *infinitary*. Section 1.1.2 focuses on (b), and attempts to explicate the notion of ‘infinitary object.’ Kripke argues that if meanings are indeed infinitary, then there is considerable difficulty in conceptualizing and explaining how they could be, in any relevant sense, things in the speaker’s head, or more generally, facts about the speaker. If, on the other hand, a putative meaning fact fails to be infinitary, it will not succeed in *constituting* my meaning anything particular by a symbol. With Section 1.2, I pass to a presentation of Goodman’s new riddle of induction. 1.2.1 explains how the riddle makes essential play with the invention of predicates with “gerrymandered” extensions, and how it exploits the fact that although we have no difficulty conceiving and manipulating such extensions, there is something about these predicates, the complexity of their *meaning*, which rules them out from use in inductive inference. In 1.2.2 I show that Goodman’s argument for the language-relativity of complexity applies equally well to Kripke’s own gerrymandered predicate, ‘quus’. The reason for which we find ourselves utterly incapable/unwilling to *use* such predicates must, I suggest, be the same in both cases. Now, a choice of predicate is tantamount to a choice of description, and both paradoxes highlight the fact that any given event falls under infinitely many different but extensionally correct descriptions. Both paradoxes therefore contain at their core the puzzle of multiple redescription. Section 1.3 is dedicated to showing that this puzzle is essentially the same as the problem of ‘curve-fitting' as encountered in the empirical sciences. Section 1.3.1 starts out by showing that the choice of predicate, in certain types of situation, is tantamount to a choice of curve over a set of data. The well-known problem of underdetermination of any curve by a finite amount of data is of the same type as the underdetermination of any given description by an object at a certain time: for, just as a given data set is compatible with infinitely many curves superimposed on it, a given object satisfies infinitely many true descriptions. The Section ends with the conclusion that one way to view the extant philosophical challenge is as a challenge to uncover the nature of the constraints (if any) operating on the sort of curve-fitting that is going on in the cases Kripke and Goodman were concerned with. The *desideratum*, clearly, is that these constraints be sufficient for excluding the relevant sceptical reinterpretations. Sec. 1.3.2 notes that, from a pragmatic perspective, the point of having a device like redescription in one's language is that of having a way of highlighting alternative features of one and the same object (the ones one is interested in). Now, the concepts of description and sameness with respect to X, or similarity, are of course interwoven concepts. I point out that Wittgenstein, for one, did in fact consider the problem of rule-following as one essentially involving the problem of what constitutes "going on the same way", and hence of description. The relativity of similarity, or being the same with respect to some description, is due to the possibility of multiple redescription, a theme further developed in Section 1.3.3. Both paradoxes, I argue there, trade on the fact that similarity, just like simplicity, is relative to choice of description. Consequently, it is not a coincidence that the realist reaction to both paradoxes has relied in part on some notion of *objective* similarity. At the end of the Section, I distinguish the debate whether the acquisition of various pre-established descriptive means during language acquisition is an inductive process (by which we stumble upon such objective similarities), rather than due to an innate mechanism (by which we activate innate structures and 'similarity spaces'), from what is in fact at stake in both paradoxes: the philosophical consequences of the brute fact that we do make choices of description that are, more or less, *compulsory*.

1.1 Kripke's "Sceptical Problem"

The main philosophical battlefield in connection with "Kripke's Paradox" is the truth of its first premise, i.e. the claim that there are no meaning-constituting facts. To establish this premise, Kripke's imaginary Sceptic proceeds by elimination, arguing that all possible candidates for being a meaning-constitutive fact ought to be rejected, because they fail either of two criteria: meaning is *normative*, and meaning is also *infinitary*. Now, the study of the connection, if any, between meaning and normativity is a mainstay of contemporary philosophy of language—however, it does not play a major role here. All focus will be on the notion of 'infinitary object', for it is this second aspect of Kripke's attack on meaning-facts that allows a *rapprochement* with Goodman. Loosely characterized, infinitary objects are objects that uniquely determine an infinite object. The meaning theorist, or rather a meaning theorist of a certain stripe, wishing to refute the paradox appears saddled on a dilemma: if he acknowledges meanings as infinitary, then it is quite mysterious how they could be, in any relevant sense, *ours*, i.e. how our minds could possibly grasp them, manipulate them, use them. If, however, he does not conceive them as infinitary, then meanings are not able to uniquely determine extensions (which may be infinite) and thus appear inadequate for the kind of job the theorist needs them to do to dissolve the puzzle.

1.1.1 Plus *vs.* Quus

Stripped down to its essentials, Kripke's presentation of what he himself refers to as 'Wittgenstein's sceptical problem' may be summed up as follows: take a mathematical function called 'quus' and denoted by the symbol '$\oplus$', such that

$$x \oplus y = \{\frac{x + y \text{ for all } x, y < 57}{5 \text{ for all } x, y \geq 57}\} \quad \text{(Kripke 1982, p. 9)},$$

and assume for the sake of argument that you have never in your life computed '68+57=125'. (Nothing in what follows will depend on the particular number at which *quus* is defined as deviating from *plus*. However, it will often be more intuitive to suppose it is a very large number). Now imagine a Sceptic who asks you to add 68 and 57. The Sceptic holds that in responding in one way rather than another, e.g. by uttering "125", you will respond without any *justification* whatsoever. This is not because arithmetic is a particularly erratic and unjustifiable business. Rather, it is because there is no *fact of the matter*, according to the Sceptic, about your meaning one thing rather than another thing by the symbol '+'. In particular, there is no fact of the matter, a discovery of which is apt to falsify the perverse suggestion that with '+' you have always meant the function *quus*, and that therefore you should now respond with '5' if you want to stay true to your past intentions. You are, it seems, in quite a predicament: if you cannot establish what you have meant by '+' in the past, then you also lack the resources to disprove the claim that you are referring to *quus* at this very moment.

Now, computing a mathematical function for previously unencountered arguments is quite similar to applying a word to a new case—a similarity Wittgenstein made much of in the *Investigations*.[14] The Sceptical challenge hence appears to concern not only our ability to mean something by arithmetical symbols, but our meaning something by any given symbol. For, just as an arithmetical function takes two natural numbers as arguments and yields another as value, the meaning of a word could, and has been, conceived as a function that takes particular situations of use as arguments and yields either a "yes" or a "no", correct or incorrect, as an answer. On this objectivist, quasi-Platonist, view, the role of linguistic meaning is construed as determining, for any one of indefinitely many possible linguistic contexts, whether or not a word has been correctly applied on that occasion (cf. Collins 1992, p. 74.). It is, incidentally, a conception of meaning that Wittgenstein was very interested in. He often passes seamlessly from discussing arithmetical symbols and their use, to natural language words and their use, and clearly endorses the analogy between the mathematical and the linguistic case (although he does not endorse the "Platonist" conception of meaning).

If the Sceptic is right, 'the entire idea of meaning vanishes into the air,' according to Kripke, a conclusion unacceptable enough to qualify as paradoxical (Kripke 1982, p. 21). The paradox may be put into the following form:

[14] Cf. Wittgenstein, L. (1967). *Philosophical Investigations. Translated [from the German] by G.E.M. Anscombe*, Oxford, Blackwell, e.g. §§ 226-41. Cf. also Kripke *Wittgenstein on Rules and Private Language*, p. 19.

(1) There are no past or present facts differentiating* between a speaker's meaning *quus* and his meaning *plus*
(2) If (1), then there are no facts about a speaker's meaning something determinate
∴ (3) There are no facts about a speaker's meaning something determinate
(4) If (3), then there is no such thing (event, state) as a speaker's meaning something determinate
∴ (5) There is no such thing (event, state) as a speaker's meaning something determinate
(6) If (5), then there is no meaning
∴ There is no meaning

* 'to differentiate', here, is taken to mean something like the following: a fact *p*, the obtaining of which is necessary for the truth of proposition P, differentiates between facts *x* and *y*, the obtaining of which is necessary for the truth of propositions X and Y*,* if and only if the truth of P entails either the truth of X or the truth of Y, but not both.[15]

Several immediate *addenda* seem required. (a) One may reject premise (6), either by invoking a distinction between meaning and speaker's meaning or (what may amount to the same), by claiming that the idea that a speaker's meaning is always indeterminate does not undermine the idea that there are meanings. This sort of objection would not be fatal, because (5) on its own is arguably unacceptable enough to let the paradox—or a closely related paradox—survive. Premise (6) simply accounts for Kripke's dramatic remark that 'the whole idea of meaning vanishes into the air'. (b) It could be argued that the above argument cannot be called a paradox, because a proper paradox is supposed to be a valid argument from acceptable premises to an unacceptable conclusion, and premise (1) is too contentious to count as a generally acceptable premise. Rather, the objection continues, (1) looks like the paradoxical *conclusion* of Kripke's discussion, which, after all, is for a large part dedicated to showing why no candidate meaning-fact could satisfy the sceptic's demand, and hence to vindicating premise (1). I concede that the substance of the discussion in Kripke 1982 concerns the suitability of various types of facts that could falsify (1), and that the above 'para-

[15] The notion of 'fact' will be a central one throughout this thesis, yet it is not a thesis *about* (the nature of) facts. In accordance with Kripke, who is not explicit about this, I shall employ a rather flat-footed notion of fact, as something that makes potential truth-bearers, such as propositions, sentences, and beliefs, true or false. It is hoped that the immediate difficulties that follow—namely whether this theory commits one to accepting negative facts, possible facts, conditional facts, facts that somehow mirror the logical structure, if any, of truth-bearers, as well as the more general problem whether there can be any such truth-making facts in the first place—although not *irrelevant* to our main concerns, have no impact on the conclusions we shall draw.

dox' is a corollary likely to follow from whatever argument succeeds in establishing (1). Moreover, a large part of this thesis will equally be concerned with candidate meaning-facts and worries concerning their eligibility—worries that can be seen as supporting premises in favour of (1).

My reasons for nevertheless construing Kripke's paradox as *departing* from, rather than concluding with, the first premise, lead me to my next comment: (c) Some critics of Kripke (and defenders of Wittgenstein) accuse the Kripkean sceptic of unwarranted reductionism about meaning. After all, why should my meaning *plus* by "+" be constituted by *other,* independently specifiable, facts? Why are we not allowed to reply, in a minimalistic vein:

> The fact that makes 'A means plus by "+"' true (the fact expressed by 'A means plus by "+"') consists in the fact that A means plus by "+"

It is the obtaining of this fact, we might say, which distinguishes between A's meaning *plus* and his meaning *quus*, and why should not appeal to this irreducible fact be sufficient to justify A's answers to arithmetical problems? Kripke acknowledges that declaring meaning and the corresponding meaning-facts to be *sui generis*, i.e. holding that they resist decomposition into further facts about the speaker, is one way out of the paradox. For example, one might think that 'A means *plus* by "+"' denotes A's own, irreducible *experience*—the state of meaning something by a symbol might be an experience as unique and irreducible as that of seeing a particular colour, or of having a particular kind of headache. If the requested meaning-fact concerns a mental *quale* of this type, then it is no wonder that it cannot be further reduced!

Kripke is rather uncharitable with this solution, declaring it off target (Kripke 1982, pp. 41-42): how could such a fact, albeit irreducible, possibly *justify* my solving the addition problem as I do, he asks, and how could it distinguish between *plus* and *quus*, and infinitely many alternatives? If meaning *plus* rather than *quus* by '+' is on the same level as having *this* rather than *that* kind of headache, then this might settle the matter for me internally, but I could not use this circumstance in order to justify to anyone else the way in which I proceed. Indeed—and this bears significant similarities to Wittgenstein's famous discussion of private language—it would seem that I could not even use this circumstance in order to justify *to myself* the way in which I proceed. How do I know that my headache today is of the *plus* or the *quus*-variety? I could be wrong... there aren't any reliable criteria of identity for types of headache, beyond those which are equally accessible to outsiders: "He's holding his hand to the right side of the head today"; or, "his MRI-pattern today is the same as yesterday's". Obviously, these criteria just are not fine-grained enough to distinguish between meaning *plus* and meaning *quus*. (If magnetic resonance images were correlated one-to-one with thoughts, the monitor on which we display them would have to have infinitely many pixels. The same goes for any other brain imaging technique, whether actual or yet to be invented).

The option of saying that the fact the Sceptic demands is irreducible and *sui generis* has been taken by some authors, who accept (1) but deny (2) (e.g. McGinn 1984),[16] and argue that the Sceptic gets off to a wrong start by misunderstanding the type of 'fact' we are dealing with. I see it as a virtue of presenting the paradox in the way I have done above that it renders salient these alternative, non-reductionist, options for rejecting the argument. A *naturalist* about meaning, however, will refuse to accept (1).[17] Whether one chooses to see premises (1)-(6) as constituting the real paradox or merely as the insignificant aftermath of the real action, which takes place when one tries to support or deny (1), shall be irrelevant to the purposes of this thesis. For our present concern is the connection between Kripke's paradox and Goodman's, and its implications for prospective solutions. We shall inquire, in particular, whether our suspicion, mooted above, is correct that naturalistic solutions would be particularly affected by a connection between the paradoxes. The question of the truth of premise (1) and the availability of specifically *naturalistic* meaning-facts (facts accessible to one of the natural sciences) will thus be a focal point of this study.

This leads me to my last comment. (d) Contrary to what some commentators appear to have thought, the sceptical challenge is directed not merely at the epistemic status of first-person authority and self-knowledge. True, it invokes the possibility of a mismatch, undetected by the subject, between the intentional object of her past intentional state *X* and the object of her present intentional state *Y*. Moreover, the dialectic setting in which Kripke presents the paradox, namely an encounter with an imaginary Sceptic about my individual meaning, fosters the impression that what is at issue here is how anyone can know what one means (Kripke 1982, p. 8*sq*). However, the Sceptic plainly asks what *makes it the case* that a subject—any subject—means this rather than that, i.e. he asks the metaphysical question of the constitution and

[16] For a defence of non-reductionism about meaning, see Boghossian "The Rule-Following Considerations", pp. 540-49.

[17] I shall use the term 'naturalist' (about meaning) to denote all those philosophers who believe that facts about meaning and other intentional mental states are just as accessible to the "natural" sciences as any other fact. A 'physicalist' about meaning will be one who believes that these facts are describable in the language of physics.

conditions of individuation of meaning itself. In fact, Kripke rules out an exclusively epistemic intepretation of the sceptical challenge: '... merely epistemological scepticism is not in question. The sceptic does not argue that our own limitations of access to the facts prevent us from knowing something hidden. He claims that an omniscient being, with access to all available facts, still would not find any fact that differentiates between the plus and the quus hypotheses.' (Kripke 1982, p. 39)[18] The reason for which we cannot know what anyone means is not that meanings are epistemically inaccessible in the same way as, perhaps, some mathematical truths are inaccessible, or in the way in which I am in certain respects opaque to myself. Rather, there *are* no facts about meanings, and this is plainly why we cannot know them. The Sceptic says that the problem of determining individual meanings is intractable—as intractable as, say, the problem of finding the largest prime number.

Kripke explicitly admits *any* type of fact as a possible candidate for meaning constitution (Kripke 1982, p. 39). In particular, he admits facts that are not immediately and exclusively accessible to the individual speaker, allowing both the first-person as well as the third-person view. This means that if the Sceptic were to be vindicated, we would need to conclude generally that not only individuals, but the whole language community—and, in particular, its scientific sub-community—are unable to distinguish between someone's meaning *plus* rather than *quus*. As Boghossian 1989, p. 508, points out, Kripke argues in favour of this the Sceptical thesis by *elimination*, by considering and rejecting one after another plausible candidates for the kind of fact susceptible of fixing the meaning of a symbol. The types of facts he explicitly considers are facts concerning

1) my (past and present) use of the symbol '+'
2) my rules (instructions, algorithms) regarding the correct application of the symbol '+'
3) my disposition(s) with regard to the symbol '+'
4) my (or someone else's) simplest hypothesis concerning the meaning of the symbol '+'
5) my "qualitative mental history" (*qualia*) associated with the symbol '+' [19]

[18] For a forceful and, in my view, definitive rejection of any epistemological interpretations of the Sceptic, see Horwich "Meaning, Use and Truth"; also Forbes, G. (1984). "Scepticism and Semantic Knowledge" *Proceedings of the Aristotelian Society* **84**: 223-240.

[19] The passages where Kripke discusses each kind of fact are Kripke *Wittgenstein on Rules and Private Language*, pp. 13-14, 15-18, 22-37, 38-39, and 41-51, respectively. For discussion of this point, see Boghossian "The Rule-Following Considerations", p. 508sq; and Stegmüller *Hauptströmungen der Gegenwartsphilosophie (Vol. 4)*, pp. 23-69. Boghossian, incidentally, has only facts of type (1), (3), and (5) in his list, perhaps because he thinks that Kripke does not consider (2) and (4) to be serious contenders. However, Kripke does discuss in several places the idea of explicit rules guiding our appli-

This list conspicuously omits a very important class of facts, i.e. facts about the neurophysiology of my brain. The explanation for this is that Kripke apparently considers these to be subsumed under facts about my dispositions. He extensively discusses Wittgenstein's enigmatic remarks about mathematical functions that are "embodied" (in other words, *realized*) in a physical machine as a variant of the dispositional approach to the mind (Kripke 1982, pp. 33-36). If one took the view that the dispositional properties of an object must ultimately be grounded in, or reducible to, a set of non-dispositional properties of the object's microstructure, then neurophysiological facts about the brain would find a natural place within a dispositional theory of meaning. The question of dispositional theories of meaning, their ontology and etiology, and potential for resolving the paradox will be discussed extensively in Chapter 2.

We follow Boghossian in concluding that Kripke rejects alls candidate facts for what are eventually one or both of the following two reasons: (a) meaning is *normative*—if I mean *plus* by '+', then this state of affairs creates truths about how I *ought* to apply a given expression, not truths about how I *will* apply it; (b) meaning has what Boghossian calls an *infinitary* character—if I mean *plus* by '+', then the number of truths about how I ought to apply the term is infinite (Boghossian 1989, p. 509). For example, if I am to use '+' in a way consistent with it meaning *plus* (rather than anything else) I should apply '+' to just the members of the set of ordered triples that is the *plus* function. However, it seems that as a finite being, no sets of facts about me could be constitutive of my mental state of meaning *plus*. Boghossian reads Kripke as actually deploying these two "tests" only against the dispositionalist response, although Kripke's stated reasons against the other kinds of fact are also based on either the argument from normativity, or that from the infinitary character of meaning. For instance, the reason for which Kripke deems a mental *quale* insufficient is that it contains no normative force, and also fails to individuate *plus* from its infinitely many quus-like cousins (see our discussion of headaches, *supra*).[20]

Similarly in the case of facts of type (2), which are an interesting category. Suppose I have explicit rules or instructions, that I have either given myself or received from members of my linguistic community, regarding the use of '+'. The twin

cation of symbols, as well as suggestions from the philosophy of science (cf. Stegmüller *Hauptströmungen der Gegenwartsphilosophie (Vol. 4)*, for more details).

[20] Kripke *Wittgenstein on Rules and Private Language*, pp. 41-51. Kripke does take it as immediately obvious that actual use, whether individual or the use of the whole community, underdetermines *plus*, and looses no time arguing for this (cf. Kripke *Wittgenstein on Rules and Private Language*, pp. 41-42).

requirements of normative and constitutive force mean that the relevant rules would have to be such that they, (a) would allow me to recognize in every instance whether I have used ‘+’ correctly, and (b) allow me to distinguish between instances of following *plus* rather than *quus*. If our explicit rules are to satisfy these requirements, then it would seem that we need to state second-order rules for how to correct apply our first-order rules. To see why this is so, Kripke considers the case of the mathematical sophisticate, who points out to the Sceptic that we can exclude the quus-function as an acceptable interpretation of ‘+’ because it does not satisfy the laws (=rules) that methematicians commonly accept for ‘+’ (Kripke 1982, p. 16n). In particular, he points out, addition, but not quaddition, is the only function on the natural numbers that satisfies the following recursion equations for ‘+’:

$\forall(x)(x+0=x)$
$\forall(x)\forall(y)(x+y'=(x+y)')$ (Kripke 1982, p. 17n)

where ‘ ′ ’ indicates the mathematically fundamental idea of ‘successor’. Why can we not say that these are these laws are our rules for the correct application of ‘+’, and that they fulfil the Sceptic’s normative and constitutive requirements? (The meaning of ‘+’ would thus be “given” by the above equations.)

This is no good, says Kripke, because the signs we have used in stating the rule for ‘+’ have all been applied a finite number of times themselves, and the Sceptic can reinterpret them in a perverse way just as he has reinterpreted ‘+’. For example, the universal quantifer $\forall(x)$ might mean ‘$\forall(x)$ for every x<h OR ... for every x>h’, where h denotes a limit to the instances in which we have hitherto applied universal instantiation, and where you can insert anything you like into the second disjunct (cf. Kripke 1982, p. 17n). If finding the relevant rules for the application of symbols is our recipe for responding to the Sceptical challenge, then we would need to state second-order laws or rules for every symbol used in the recursion equations in order to rule out the Sceptic’s disingenious reinterpretations. But now it is obvious that the Sceptic could play his game at that level again, so that we would need third-order rules for the application of the second-order rules, etc. *ad infinitum*.[21] The same is

[21] We shall encounter that very same strategy in a different guise when discussing Goodman’s ‘grue’. At moments such as these (i.e. Kripke *Wittgenstein on Rules and Private Language*, pp. 15-17), when he expounds what is known as the infinite-regress-of-rules argument, Kripke comes closest to Wittgenstein’s original remarks on rule-following: a rule (or rather, the ‘expression’ of a rule), Wittgenstein famously held, is simply unable in and of itself to uniquely determine an action, in the sense of incorporating instructions concerning its own correct application. For any rule can be *interpreted*, or understood, in myriad ways. This is Wittgenstein’s original “paradox”, stated in Wittgenstein *Philoso-*

true, Kripke points out, if we tried to state our rules for '+' not in a mathematically sophisticated way, but in terms of a set of practical instructions for how to perform addition by manipulating piles of marbles (Ibid.). (Obviously, the idea of 'successor' would be a fundamental one in this set of instructions, as well).

I shall now leave the detail of the Kripkean dialectic, of specific arguments and counter-arguments, for my purpose is not to engage in a detailed defence of Kripke's paradox, but to look at it a different degree of resolution. From a greater distance, it strikes the eye that a large part of the literature on Kripke's paradox has centred on wrestling with his claims concerning the normativity of meaning-facts, rather than on their infinitary character. Kripke deploys the argument from normativity to good effect against many kinds of putative meaning-fact that have been thrown into the ring against the Sceptic. As a response, several authors have devised accounts of linguistic normativity to show how the latter is either entirely compatible with the facticity of meaning, or that linguistic meaning is not intrinsically normative at all.[22] The question of normativity, however, is orthogonal to our present concerns—a claim I cannot fully motivate at this stage of the discussion, but which I hope shall become plausible as we proceed. In a departure from the literature, then, I shall focus on Kripke's paradox as a riddle for those who conceive meaning to be both reducible to matters of fact, and infinitary. For it it is precisely the difficulty raised by the purported infinitary character of meaning, that unveils the connection between Kripke's paradox and the riddle of induction, and shows both problem complexes to be two sides of the same coin of a wider philosophical conundrum.

---

phical Investigations, §202. The consequence he draws there is that the act of following a rule must be a practice, rather than any sort of interpretation of the rule. Although Wittgenstein is, within Anglo-American analytic philosophy, widely credited with this 'rule-following argument', it can in fact be traced back in this very form to Kant, who wrote in the *Critique of Pure Reason*: 'If judgement wanted to show universally how one is to subsume under these rules, i.e. distinguish whether something belongs under the rule or not, this could only happen via a further rule. But because this is a rule it requires once more an instruction by judgement, and thus it is shown to be the case that the understanding is admittedly capable of being instructed and equipped by rules, but that judgement is a particular talent which cannot be given by instruction but can only be practised.' (Kant, I. (1787/1983). *Kritik der reinen Vernunft (Erster Teil)*, Darmstadt, Wissenschaftliche Buchgesellschaft, B, 172. Quoted in Bowie, A. (1999). "The Meaning of the Hermeneutic Tradition in Contemporary Philosophy" *German Philosophy Since Kant*. A. O'Hear, London, p. 126.

[22] Examples of the first strategy are Fodor, J. A. (1987). *Psychosemantics*, Cambridge, Massachusetts, MIT Press, Fodor "A Theory of Content II", and Millikan "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox". An example of the latter is Horwich, P. (1998). *Meaning*, Oxford, Clarendon.

1.1.2 'Infinitary' Meaning

Any 'infinitary object' in Boghossian's sense is an object which uniquely determines an infinite object. Take concepts, for instance. It is commonly considered a necessary and sufficient condition for an entity's membership in the set of green things that the concept GREEN applies to it. Correspondingly, for every entity X both actual as well as possible, there will be a true proposition asserting or denying that X is subsumed under GREEN. The concept GREEN in this sense "picks out", or is in one-to-one correspondence with, the set of actual as well as possible green things, a set likely to be infinite. Concepts are thus infinitary objects, along with functions, intensions, properties, and universals. By contrast, individual rocks, chairs, Peter and Mary, *this* man over there, or *that* configuration of neurons, are not infinitary. Although infinitely many propositions may be true *of* them, they themselves do not generate truths in the way concepts, properties, functions, etc. do, and they do not uniquely determine infinite objects. For example, given two integers as arguments, neither Peter nor *this* neural network yields a third integer as a value in the way the addition function does. Being infinitary is not the same as being infinite, however, or as having an infinite extension. Concepts whose extension is patently finite are nevertheless infinitary: the set of first Presidents of the United States, for instance, contains only one member. The extension of the concept FIRST PRESIDENT OF THE U.S is thus {George Washington}. Nevertheless, the concept, *qua* concept, generates truths about its correct application in infinitely many situations, actual and possible. Suppose, for example, that I had dreamt about George Washington (which I have never done). Then, if I possess the concept FIRST PRESIDENT OF THE U.S., this engenders a truth about a necessary condition for being a correct report of my dream. For every entity in the universe, whether actual or possible, the concept, just like a function, yields a truth about whether it applies to it. Although usually construed as finite entities themselves, concepts can "do" something infinite.

If this analysis of concepts is nearly correct we seem to have stumbled upon a dilemma: given that concepts are indeed infinitary in the sense described, then it is difficult to see how they could ever be *ours*. The Kripkean Sceptic argues that no known fact or collection of facts about any of our mental states, whether occurrent or dispositional, qualitative, historic, etc., could be such as to *constitute* the alleged infinitary object.[23] Concepts *qua* infinitary resemble Gottlob Frege's *Begriffe*, i.e. abstract "unsaturated" entities (see e.g. Frege 1994, p. 22). Any such entity immediately gives rise to the difficulty of accounting for our way of grasping it, or entering into any cognitive relation with it. On the other hand, if we prefer to think that concepts are not abstract, but rather concrete mental particulars—e.g. 'exemplars', 'prototypes', or other kinds of mental representations hypothesized by cognitive psychologists—then, and this is the second horn of the dilemma, it will be difficult to explain how concepts could possibly perform their infinitary job of uniquely determining the set of things that fall under them. After all, exemplars, stereotypes, etc. are mental representations realized in the finite brain of a cognitive agent with finite cognitive powers. As such, they are finite entities located in a finited space with a finite life time (imagine they are, for instance, intricate patterns of neural network activations). The obvious option at this point would be, of course, to avoid the dilemma by abandoning as too strong the demand—customary among many analytic philosophers (Frege first and foremost), but not psychologists—that concepts uniquely determine the set of things falling under them.[24] Concepts, on this alternative view, perform no such things as we have just attributed to infinitary objects. The trouble is that such more "down-to-earth" concepts could not be appealed to in a reductive argument against the Sceptic, for the latter demands that my meaning *plus* be reducible to some mental fact about me that individuates my meaning by distinguishing it from any one of the infinitely many alternatives. Concepts *qua* concrete mental representations cannot do that, or so it seems.

Almost the same considerations as for concepts apply to 'meanings.' The meaning of a word will be infinitary if we grant the view that word meaning determines, for any of an indeterminately large number of things, whether the word applies to it. If meanings fulfil this role, then it is difficult to understand how we could ever grasp them, know them, and manipulate them—how these meanings could be

---

[23] He argues for the same conclusion with respect to dispositional mental states, and "qualitative" ones, at Kripke *Wittgenstein on Rules and Private Language*, pp. 26-28 and pp. 41-51, respectively.

[24] Cf. the discussion in Margolis, E. and S. Laurence, Eds.) (1999). *Concepts: Core Readings*, Cambridge, Massachusetts, MIT Press, Introduction. In fact, psychologists do seem in their majority to view concepts as particular representations quite literally in the head of cognitive agents that perform no such miraculous feats as philosophers ascribe to them. Jackman, H. (2000). "Foundationalism, Coherentism and Rule-Following Skepticism", p. 13, points out that studies of the psychology of classification seem to indicate that we do *not* conceptualize our experiences in terms of categories determined by sets of necessary and sufficient conditions, for we often lack any firm disposition to place objects or situations either within or outside the extension of a given term (cf. also the well-known study Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*, Chicago, Univ of Chicago Pr).

ours. Take the meaning of *plus*, or that of the mathematical symbol '+'. For every pair of numbers, whether familiar or hitherto unencountered, it ought, at least on the infinitary construal, generate a truth about what constitutes a correct application of '+' to any of these numbers. If so, how am I supposed to *access* that truth, Wittgenstein famously asked: the meaning of '+' does not seem to supply me with "instructions" towards its correct application for every new case. *Ditto* for colour words: how do I know what 'red' means, in the sense of how do I know how to correctly apply it to *this* new thing that I have never seen before? When I learned the meaning of 'red' I was certainly not given *explicit*, nor indeed any implicit, instructions as to what to do in new cases (e.g. Wittgenstein 1967, §§239, 273-74). Nevertheless, we all sometimes have the impression that (and do speak as if) we were being "guided" by the meaning of words—like a railway car engaged on a set of rails, according to Wittgenstein's metaphor. One of the perhaps best assimilated lessons of the *Investigations* is precisely his warning not to confuse the phenomenal quality of the *Bedeutungserlebnis* ("meaning experience") associated with a word, such as being guided by it in a peculiar way, with the meaning of a word *per se*.

The Fregean approach to this set of problems is the following: the word 'plus' has a *sense,* which uniquely determines its *reference*. The sense of a linguistic expression contains all that is necessary and sufficient to determine the thing(s) to which the expression truthfully applies. This view is sometimes called 'Meaning Platonism,' for it entails that the sense of an expression determines *user-independently* all future and possible correct applications of that expression (its extension) (Puhl 1991, p. 1). When we 'know the meaning' of a word, or know the unique arithmetic rule expressed by a symbol such as '+', we somehow 'grasp' or are otherwise in cognitive contact with its sense. It is this cognitive relation, which presumably allows us to determine for any situation and any entity whether a given expression would correctly be applied to that entity in that situation. Without senses, and our ability to access them, communication would be impossible, Frege believed, for otherwise we could not be assured to be talking and thinking about the same thing as our interlocutors (Frege 1892). Thus, it is our grasp of the sense of '+' in virtue of which we are capable of ascertaining whether the assertion "68+57=125" constitutes or not a correct application of '+' to the numbers 68 and 57. And it is our grasp of the senses of linguistic items that should, on this view, function as the court of appeal for sceptical challenges such as the above.

Now, Kripke acknowledges that it is in the *nature* of senses—if indeed there are such things—to determine referents, but he maintains that even so

> … the sceptical problem cannot be evaded, and it arises precisely in the question how the existence in my mind of any mental entity or idea can constitute 'grasping' any particular sense rather than another. The idea in my mind is a finite object: can it not be interpreted as determining a quus function, rather than a plus function? Of course there may be another idea in my mind, which is supposed to constitute its act of *assigning* a particular interpretation to the first idea; but then the problem obviously arises again at this level. … Platonism is largely an unhelpful evasion of the problem of how our finite minds can give rules that are supposed to apply to an infinity of cases. Platonic objects may be self-interpreting, or rather, they may need no interpretation; but ultimately there must be some mental entity involved that raises the sceptical problem. (Kripke 1982, p. 54)

Even if we allowed that "Platonic entities" are such that they achieve the feat of being self-interpreting—in the sense of telling us, in and of themselves, how they are to be correctly manipulated and applied on all occasions—it remains a moot point how something *mental* and hence finite could possibly perform the same sort of job. In Boghossian's terminology, Kripke's Sceptic claims that although Platonic entities may indeed be infinitary, nothing mental could be. If we grant this, however, we seem unable to explain how speakers can make cognitive contact with an infinitary entity and, in particular, how they can know that they have grasped one and not the other. As David Pears puts it: '… Platonism makes an impossible demand on the word-user's mind: it is required to contain something which is both strictly contemporary and also a self-contained, unambiguous representation of the infinite line dividing positive from negative instances. The impossibility of meeting this demand shows up in the infinite regress that it generates.' (Pears 1988, p. 11).[25]

Why is it, precisely, that the word-user's mind could not contain something infinitary? Evidently, a finite mind cannot *literally* contain something that is infinite, for example a representation of every member of the set of ordered triples constitutive of the addition function—the extension of '+'. Kripke holds that it is equally difficult to accept that the mind could contain something that is "strictly contemporary", i.e. finite, but that nevertheless distinguishes *plus* from *quus* by somehow doing duty on an infinite number of occasions. According to Boghossian,

> The subtle point that Kripke relies on here is that it is as hard to explain how a finite mind might grasp an infinite object—such as the addition table—

[25] Pears alludes to the infinite regress we have mentioned above, i.e. Wittgenstein's original rule-following argument in Wittgenstein *Philosophical Investigations*, § 201 and preceding.

> directly, as it is to explain how it might grasp something that uniquely determines such an infinite object. For in the relevant sense, an object that uniquely determines an infinite object itself has an infinite number of individuative conditions, and hence is itself an infinite object. If, then, there is a problem about grasping infinite extensions, that problem simply resurfaces for senses. (Boghossian 1994, p. 140)[26]

In other words, infinitary objects must themselves be in some respect infinite. This passage, in my experience, often causes puzzlement: what exactly does Boghossian mean by the term 'individuative conditions'? It would be correct, but hardly illuminating, to say that an infinitary object's 'individuative conditions' are precisely those properties, whatever they are, that allow it to uniquely determine an infinite object—for it is of course the having of such properties (if any) that make an object infinitary. The comparative obscurity of the notion of an object's having infinitely many individuative conditions stems from the obscurity of the notion of an 'object that uniquely determines an infinite object', i.e. from the notion of an infinitary object itself. Whether we choose one characterization or the other, we scarcely leave this murky zone. Perhaps the sceptic's demand for facts establishing the infinitary role of meaning cannot be satisfied from the start, because it is founded on a misunderstanding—perhaps meaning simply is not infinitary?

Here we have, of course, come across one of the many reasons for which Kripke's Wittgenstein is difficult to reconcile with what is usually thought to be the actual Wittgenstein's position. Wittgenstein, like many other philosophers, found precisely the infinitary conception of meaning entirely unintelligible. In the *Investigations*, he frequently points out to his fictitious interlocutor (who is, at times, quite distinctly Fregean) that meaning something by a given expression, for example meaning something by the order '+2', *just could not be* determining in advance its extension, the sequence of even natural numbers. Meaning something by a word is not being connected to something that determines, user-independently, where the train is going. Wittgenstein describes his interlocutor's infinitary approach to meaning in an ironical vein: 'Your idea was that that act of meaning the order [to keep adding 2] had in its own way already traversed all those steps: that when you meant it your mind as it were flew ahead and took all the steps before you physically arrived at this or that one. … And it seemed as if they were in some unique way predetermined, anticipated—as only the act of meaning can anticipate reality.' (Wittgenstein 1967, §188).

---

[26] For a similar point, see also Stegmüller *Hauptströmungen der Gegenwartsphilosophie (Vol. 4)*, pp. 70-73.

Regarding the case of colour words, which Wittgenstein used interchangeably with the arithmetic example, David Pears comments:

> If our use of colour words were not guided by rails laid down in reality, the requisite conventions would be [completely] arbitrary. Or so the Platonist claims. Wittgenstein objects that this dilemma is founded on an illusion. It seems to us that we have to choose between Platonism and pure conventionalism only because we have allowed ourselves to be persuaded that Platonism really would provide us with a satisfactory explanation of the stability of our colour language, if we could believe it. But would it? ... the trouble with Platonism is not only that its fixed rails are a fantasy. Even if they were palpable, we could not be guided by them unless our minds received self-contained, unambiguous representations of their infinite continuations. But that is not possible. Or so Wittgenstein claims. (Pears 1988, pp. 14-15)

In the vocabulary of Kripke's sceptical paradox, this translates to: there are no facts in the world about what a given word means (= no fixed rails), and there are also no facts about items in my head that uniquely determine what I mean by that word (= no representations of the fixed rails in our minds). If this was the case, the sceptic would win by default; but rather than having shown that there is no meaning, his achievement, in Wittgenstein's eyes, would be limited to having made it painfully obvious that he has a faulty conception of meaning in the first place.

The important issue for our present purposes is not how Wittgenstein would have dealt with Kripke's Sceptic, but rather how certain other philosophers have done. Some have seen no insurmountable difficulties involved in refuting Kripke's arguments from finitude (in contrast to his argument from normativity, which is ordinarily considered a much more serious obstacle to a naturalistic account of meaning). Theorists espousing realism and/or naturalism about meaning, in particular, have devised accounts to show how there can be something contained in my mind which, although finite, leaves no room for being interpreted as determining a *quus* function, rather than a *plus* function. In essence, these authors accept the infinitary conception of meaning, and attempt to show how such a conception escapes the sceptic's nihilistic conclusions. Representative examples of this kind of theory will take centre stage in the second and third chapters of this study.

For now I conclude my discussion of Kripke's paradox. It concentrated on the infinitary character of meaning, and the problem of explaining how finite beings could grasp, understand, distinguish, achieve any cognitive contact with, an infinite entity. The main sceptical problem has been canvassed, namely how we can assure ourselves that the process of going from the former to the latter is uniquely determined. How can my mental state of meaning *plus* be correlated one-to-one with the relevant infinite set of ordered triples? Undoubtedly, the list of issues invoked by the paradox could be extended, and there is scope for arguing that other aspects of it are least equally, if not more important—notably the normativity issue. But the purpose of this Section was merely to bring into clearer focus some of the issues that are congenial to Goodman's paradox. The exact nature of the affinity shall become clearer in the next section.

1.2 Goodman's Riddle of Induction

Goodman's 'new riddle of induction' makes essential play with predicates with gerrymandered extensions. It exploits the fact that although we have no problems with conceiving and manipulating such extensions, there is something about gerrymandered predicates, namely the complexity of their meaning, which rules them out from use in inductive inference. The language-relativity of simplicity, emphasized by Goodman in order to rebut Carnap's objections, also shows that there is nothing intrisically complex about Kripke's own gerrymandered predicate, 'quus'. The upshot of this observation is that the reason for which we find ourselves incapable/unwilling to use such predicates could be the same in both cases. This reason is, I suggest, our tendency to find 'naturally compelling' some predicates rather than others. A choice of predicate, however, is a choice of description. Any given event thus falls under infinitely many different but extensionally correct descriptions, and both paradoxes are essentially about the (apparent absence of) reasons of preferring one such description over its infinitely many alternatives.

1.2.1 The Riddle

The Sceptic's worries about the factuality of meaning bear little initial resemblance to problems having to do with induction. Aristotle, from whom the term derives, characterised induction as the 'proceeding from particulars up to a universal.'[27] Induction has traditionally been considered a non-demonstrative reasoning in support of a general proposition, the support in question being somehow produced by the consideration of particular cases that are thought to "fall under" it. David Hume famously gives a pessimistic assessment of knowledge resulting from inductive inference by pointing out that

> If reason determin'd us, it wou'd proceed upon that principle, that instances of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same. [However] … there can be no demonstrative arguments to prove, that those instances, of which we had no experience, resemble those, of which we have had experience. (Hume 1777/1975, Book I, Part 3, Sect. 6)

Induction takes observed patterns of occurrences, or regularities, in nature and uses them to make predictions about unobserved occurrences. It thereby assumes that the relevant patterns or regularities will continue to hold in the future, because "the course of nature continues uniformly the same".

The characteristic feature of Goodman's work on induction—prompting him to declare that his riddle is the "new" problem of induction superseding Hume's—is the shift of focus from the justification of the principle of uniformity, to an analysis of the concept of *resemblance* between past and future events. Thus, Goodman writes:

> That the future of induction will be like the past is often regarded as highly dubious—an assumption necessary for science and for life but probably false, and capable of justification only with the greatest difficulty if at all. I am glad to be able to offer you something positive here. All these doubts and worries are needless. I can assure you confidently that the future will be like the past. … [But] I must add that while I am sure the future will be like the past, I am not sure in just what way it will be like the past. … The question is *how* what is predicted is like what has already been found. Along which, among countless lines of similarity, do our predictions run?' (Goodman 1972, p. 441)

[27] In the newest translation, Aristotle (1997). *Topics: Books I and VII with excerpts from related texts. Translated with a commentary by Robin Smith*, Oxford, Clarendon Press, Book I, 12, 105[a].

Elsewhere, Goodman characterises induction as 'the projection of characteristics of the past into the future, or more generally of characteristics of one realm of objects into another' (Goodman 1946, p. 383) and claims that in an inductive inference, 'within certain limitations, what is asserted to be true for the narrow universe of the evidence statements is confirmed for the whole universe of discourse.' (Goodman 1983, p. 72). This logician's way of putting things makes induction appear rather like a function that takes us from the set of things in its domain to the set of things in its range. Actual inductive practice is rather complex, however, and we have multiple options when it comes to logical form, as Frank Jackson makes clear:

> …one common inductive practice we take to be rational is to project common properties from samples to populations, to argue from certain *F*s being *G* to certain other *F*s being *G*. There are many ways we can try to spell out this practice in semi-formal terms: by saying *'Fa & Ga'* confirms '$\forall x\ [Fx \supset Gx]$,' or 'All examined *A*s are *B*' supports 'All unexamined *A*s are *B*,' or 'Fa_1 & … & Fa_n' gives good reason for 'Fa_{a+1},' and so on. (Jackson 1975, p. 113.)

Also, there is of course statistical inductive inference, where we infer from the fact that x % of observed *a*'s are *B*, that x % of all *a*'s are *B*—which will lead us to believe that there is an x % chance that the next observed *a* will be *b*. This is also called frequency-induction. Finally, there is inference 'by analogy,' where we conclude from the observation that some $b_1 \ldots b_n$ share the properties F_1, $\ldots F_n$, and the observation that *a* has $F_1 \ldots F_{n-1}$, that it also has F_n.

Goodman did not concern himself with any particular way of bringing out the logical form of inductive judgements (although one of the above schemas does appear in Goodman 1946), and we shall not do so either, heeding Goodman's point that the problem of induction, however construed, is really about 'how what is predicted is like what has already been found'. Goodman 1946 was a reaction to attempts by Rudolf Carnap, Carl Hempel, Oppenheimer, and others and others to establish a theory of *confirmation*.[28] According to these authors, certain propositions expressed by evidence statements stand in the relation of confirmation to certain other types of proposition, expressed by hypotheses or generalizations. Carnap and Hempel inter-

[28] Cf. e.g. Carnap, R. (1945). "On Inductive Logic" *Philosophy of Science* **12**: 72-97; Carnap, R. (1947). "On the Application of Inductive Logic" *Philosophy and Phenomenological Research*: 8 133-148; Hempel, C. G. (1945). "Studies in the Logic on Confirmation, Part I" *Mind* **54**: 1-26.

preted confirmation as a *logical* relation analogous to, though weaker than, that of implication, its description being the business of the inductive logic to be developed. Carnap, for instance, declared that 'inductive logic [can be] constructed out of deductive logic by the introduction of the concept of degree of confirmation.' (Carnap 1947, p. 74) Although Goodman shared Carnap's and Hempel's sense of urgency with respect to the need of finding a satisfactory theory of confirmation, he objected to the latter's positive proposals.

Goodman's worry is that no matter how sophisticated our mathematical apparatus (such as Carnap's confirmation function *c**, cf. Carnap 1945, p. 74), the predictions we obtain from it will always depend on the way the evidence has been *described*, in particular on the type of predicate used. For by manipulating a given predicate's extension it is possible to formulate two mutually inconsistent hypotheses (generalizations) both of which appear equally well confirmed by one and the same type of evidence. Elaborating on an earlier example, Goodman sets out the problem as follows:

> Suppose that all emeralds examined before a certain time *t* are green. At time *t,* then, our observations support the hypothesis that all emeralds are green.[29] … Now let me introduce another predicate less familiar than "green". It is the predicate "grue" and it applies to all things examined before *t* just in case they are green but to other things just in case they are blue. Then at time *t* we have, for every evidence statement asserting that a given emerald is green, a parallel evidence statement asserting that that emerald is grue. And the statements that emerald *a* is grue, that emerald *b* is grue, and so on, will each confirm the general hypothesis that all emeralds are grue. (Goodman 1983, pp. 72-73)

Both 'green' and 'examined before *t* and green or not examined and blue' (i.e. 'grue') syntactically are predicates, and there is no denying that the evidence statements as described in Goodman's text *support*—in the minimal sense of 'are an instance of'—both the general proposition 'All emeralds are grue' as well as the familiar 'All emeralds are green'. Clearly, for every conceivable predicate used in a general proposition and its supporting evidence statement, it is possible to find a predicate with a different extension, which is equally applicable in a description of the evidence, but instantiates a different general proposition in conflict with the first. Expressed in one of

[29] Thompson, J. J. (1966). "Grue" *Journal of Philosophy* **63**, p. 239, points out that Goodman's choice of example was unfortunate, for emeralds are *by definition* green beryls, their colour being used to distinguish them from other beryls. Although this would make 'All emeralds are green' analytic, it does not, of course, affect the philosophical significance of Goodman's argument.

Jackson's schemas, Goodman showed that for every true proposition of the form '$F(a)$ & $G(a)$' that apparently confirms the general proposition '$\forall(x)\ [F(x) \rightarrow G(x)]$', it will also be true that '$F(a)$ & $G(*a)$', apparently confirming the incompatible proposition $\forall(x)\ [F(x) \rightarrow G*(x)]$. The *crux*, for formal theories of confirmation, is that '$F(a)$ & $G*(a)$' stands in exactly the same logical or syntactic relation to '$\forall(x)\ [F(x) \rightarrow G*(x)]$ as '$F(a)$ & $G(a)$' with respect to '$\forall(x)\ [F(x) \rightarrow G(x)]$'. Putnam put the impact of this argument as follows: 'What [Goodman] proved, even he did not put it that way, is that inductive logic isn't formal in the sense that deductive logic is. The *form* of an inference, in the sense familiar from deductive logic, cannot tell one whether that inference is inductively valid.' (Goodman 1983, Preface).

We can now express the paradox in the following terms: given Goodman's definition of grue,

> $\forall x$ [x is grue iff (x is green & observed before t) $\cup$ (x is blue & $\neg$ observed before t)] (Goodman 1983, pp. 72-73.)

the following argument appears to be true.

(1) All past and present observed emeralds are green as well as grue
(2) If (1), then all past and present facts that confirm the inductive generalisation 'All emeralds are green' also confirm the incompatible* inductive generalisation 'All emeralds are grue'.
∴ (3) All past or present facts that confirm the inductive generalisation 'All emeralds are green' also confirm the inductive generalisation 'All emeralds are grue'.
∴ (4) No past or present facts can differentiate** between 'All emeralds are green' and 'All emeralds are grue'
(5) If (4), then for every inductive generalisation there is an infinite number of incompatible inductive generalisations such that no past or present facts can differentiate between them.
∴ (6) No past or present facts can confirm any inductive generalisation

(* propositions A and B are *incompatible* if they cannot both be true at the same time. ** 'to differentiate' is taken in the same sense as in our statement of Kripke's paradox (*supra*)).

Goodman acknowledges that predicates of the type of 'grue' are not, as required by the theories of confirmation developed by Carnap and Hempel, 'logically independent' from the ones they have been assembled from. But, in his eyes, all this means is that the conflicting confirmations could not occur in any one 'system'—i.e. in any one formal language, such as Carnap's language of science L. Given that 'the system containing the predicate "grue" alone is quite as admissible as the one containing "green" alone' (Goodman 1946, p. 384),[30] the requirement of logical independence does not advance us any further. 'Admissible', here, means of course admissible from the purely formal point of view. Any theory of inductive inference, even if formally admissible, will be inadequate from the *descriptive* as well as the *normative* standpoint if it countenances as "projectible" predicates with 'grue'-like extensions. (In other words, if it accepts them as predicates which may be used in valid inductive inferences.) We simply do not make any inductive inferences using such predicates, and there seems to be no particular reason why we *should*. There is, in any case, no purely syntactical or logical criterion, which affords an explanation of why we intuitively take one hypothesis to be confirmed by its instances rather than another, which is formally equally well confirmed by the same set of evidence. An *infinite* number of different propositions can be said, from a purely extensional point of view, to describe a given series of events without, however, receiving any degree of credibility from its occurrence. The problem is therefore to determine which hypothesis (or general proposition) a particular series of events "conforms to" in a way such as to constitute confirming evidence for it, above and beyond merely being described by it.

Goodman's conclusion was: 'Undoubtedly we do make predictions by projecting the patterns of the past into the future, but in selecting the patterns we project from among all those that the past exhibits, we use *practical criteria* that so far seem to have escaped discovery and formulation' (Goodman 1946, p. 385). Famously, Carnap replied that a presupposed requirement of his system of inductive logic, just as that of deductive logic, is that the properties and relations designated by its primitive predicates—those that are logically independent of one another—are unanalysable into further components, i.e. that they are *simple* (Carnap 1947, pp. 134-36)· Carnap has a distinction between purely *qualitative* properties that can be expressed using only primitive predicates, purely *positional* properties that can be expressed using only individual constants (referring, for example, to space-points or time-points), and 'mixed' properties that are neither purely qualitative nor purely positional (Carnap 1947, p. 138). The property expressed by the predicate 'grue' is, on this account, mixed and hence complex, for it has a positional component. But, Carnap stipulates, only purely qualitative, non-positional properties should be considered inductively projectible.

[30] I have adapted the quote to the later "grue" example.

The appeal to the qualitative-positional distinction is essentially an appeal to the *meanings* of 'green' and 'grue': the predicate 'grue' is such that its meaning is complex, because it can be analysed into a primitive predicate and a term referring to time; this is not the case for 'green'. Richard Jeffrey comments on the difference in philosophical outlook separating Carnap from Goodman that is illustrated in this response:

> … the permanent impact of Goodman's Query on Carnapian confirmation theory consists in having made it abundantly clear that *credibilities depend on meanings*, so that in choosing a *c* function [the function $c(h,e)$ that allows us to calculate the degree of confirmation of the sentence *h* given the evidence *e*] for a language it is essential to consider the meanings of the terms in its vocabulary. (Jeffrey 1966, p. 285; my emphasis)

In a footnote, Jeffrey adds 'Goodman, of course, would not put the matter in terms of meanings; but from Carnap's point of view, "the features of the hypothesis other than its syntactical form" on which its confirmation depends are features of the meanings of the hypothesis and of the evidence statement' (Ibid.). It is important for our topic to note that many authors after Carnap have made the same, quite intuitive, move in order to counter Goodman's arguments. *The* salient distinguishing feature between ordinary predicates and Goodman-type ones is, as most would agree, their meaning—the meaning of the former is somehow simple, whereas that of the latter is complex. (The same cannot be said of their extensions; extensions are simply sets, and all parties agree that there can be no simplicity measure for sets).

1.2.2 Simplicity Relativized

Goodman's reaction to Carnap's defence was characteristically similar to that of the later Wittgenstein to certain views expressed in his own *Tractatus*.[31] According to Goodman, the assumption that there are simple properties relies on a doubtful con-

[31] Wittgenstein, L. (1922). *Tractatus logico-philosophicus*, London, Kegan Paul, Trench, Trubner. Wittgenstein assumed the existence of absolutely simple objects (*Gegenstände*), whose arrangements are the states of affairs which constitute "everything that is the case", i.e. the world. The names of simple objects, correspondingly, are the ultimate elements into which sentences can be analyzed, and they are also absolutely simple. The concept of absolute simplicity is thus a *sine qua non* of the metaphysics of the *Tractatus*.

ception of absolute simplicity, or unanalysability: ‘The nature of this simplicity is obscure to me, since the question whether or not a given property is analyzable seems to me quite as ambiguous as the question whether a given body is in motion. I regard “unanalyzability” as meaningful only with respect to a sphere of reference and a method of analysis, while Carnap seems to regard it as having absolute meaning’ (Carnap 1947, p. 149). This is a take on the problem of simplicity very much in tune with the later Wittgenstein. Compare:

> “Simple” means: not composite. And here the point is: in what sense ‘composite’? It makes no sense at all to speak absolutely of the ‘simple parts of a chair’. ... The question “Is what you see composite?” makes good sense if it is already established what kind of complexity—that is which particular use of the word—is in question. … We use the word “composite” (and therefore the word “simple”) in an enormous number of different and differently related ways. (Wittgenstein 1967, §47)

Goodman’s belief that simplicity is always relative to a language was already implicit in his claim that a language containing ‘grue’ is, within the framework of formal confirmation theory, just as acceptable as any (see *supra*). Goodman is well-known for having artfully illustrated this point as follows:

> True enough, if we start with “blue” and “green”, then “grue” and “bleen” will be explained in terms of “blue” and “green” and a temporal term. But equally truly, if we start with “grue” and “bleen”, then “blue” and “green” will be explained in terms of “grue” and “bleen” and a temporal term. … Thus qualitativeness is an entirely relative matter and does not by itself establish any dichotomy of predicates.’ (Goodman 1983, p. 80)

Whether or not ‘grue’ appears disjunctive—i.e. both ‘positional’ and ‘qualitative’—and ‘green’ qualitative, is therefore relative to our choice of language (or ‘language-game’, as Wittgenstein might have said). Both Goodman and Wittgenstein thus hold that there is no such thing as an *intrinsically* disjunctive predicate, a predicate that would somehow be complex in all languages.

The language-relativity of simplicity obtains trivially if there is freedom in the choice of primitives, in Kripke’s case just as well as in Goodman’s. Let us change Kripke’s original definition and define the primitive terms ‘quus’ and ‘quinus’, denoted by ‘⊕’ and ‘⊗’, as follows:

$$x \oplus y = \{\frac{x + y \text{ iff } x, y < 57}{x - y \text{ iff } x, y \geq 57}\}$$
$$x \otimes y = \{\frac{x - y \text{ iff } x, y < 57}{x + y \text{ iff } x, y \geq 57}\}$$

Then, from the point of view of a language in which we take 'quus' and 'quinus' as primitives and consider them simple, our familiar 'plus' and 'minus,' denoted by '+' and '–', will appear as complex:

$$x + y = \{\frac{x \oplus y \text{ iff } x, y < 57}{x \otimes y \text{ iff } x, y \geq 57}\}$$
$$x - y = \{\frac{x \otimes y \text{ iff } x, y < 57}{x \oplus y \text{ iff } x, y \geq 57}\}$$

Of course, taking '⊕' and '⊗' as primitive arithmetic operations spells doom for any reasonably efficient and conspicuous mathematics. Attempting to do complex calculations with 'quus' and 'quinus' as basic arithmetic functions would be more than just *impractical*, it would border on the (humanly) impossible. However, given the similarities in the way in which the extensions of 'grue' and 'quus' have been constructed out of the extensions of more familiar terms, the impossibility in question is essentially the same for both terms. What is its cause? According to Christopher Peacocke,[32] a major insight of the later Wittgenstein was that '… an account of what is involved in employing one concept rather than another, following one rule rather than another, has at some point to mention what thinkers employing the concept find it natural to believe' (Peacocke 1992, p. 13). Peacocke thinks that this is precisely the reason for Wittgenstein's insistence that to grasp a particular rule, or to possess a particular concept, is to have already settled how to apply it in some (but not all) future cases. Accepting that insight, as Peacocke does, implies accepting that '… which concepts a thinker is capable of possessing depends on the ways in which he is capable of finding it natural to go on.' (Peacocke 1992, p. 14). It seems plausible that the reason why we are virtually incapable of doing mathematics with '⊕' and '⊗' as primitive arithmetic operations has nothing to do with the fact that these symbols express concepts which are (apparently) intrinsically complex, and everything to do with the 'ways in which we find it natural to go on.' Similarly, it seems that the relevant difference between 'projectible' and 'unprojectible' predicates, rather than being

[32] Peacocke, C. (1992). *A Study of Concepts*, Cambridge, MIT Pr.

the degree of complexity of their meaning, as Carnap thought, is best characterised in terms of the differences in the way a speaker using one or the other would have to 'find it natural to go on.' Clearly, if Peacocke is right concerning the role of the 'transitions or steps we find naturally compelling', Kripke's and Goodman's paradox will be rather close cousins of each other.

Goodman's use of grue-like predicates was designed to shift the focus of the problem of induction from the task of formulating and justifying inductive rules, to that of explaining why we chose to use certain predicates and not others in our actual inductive predictions. Hume's answer to the question 'Why one prediction rather than another?' was an appeal to psychological habit created by the regularities we have observed in the past—but Goodman famously quips 'regularities are where you find them!', and asks why we happen to find one series of events regular, or "naturally compelling", rather than another. This an interesting, and new, challenge: why do we happen to say and believe that 'All emeralds are green', given that all emeralds are also grue, gred, grellow, etc., and have perfectly *regularly* been so since their first discovery? We can put the point in terms of either concepts or meanings: my criteria for an acceptable inductive inference should sanction my inductive employment of colour words that express the concepts of GREENNESS, BLUENESS, etc., but exclude GRUENESS and BLEENNESS. Alternatively, we need to establish that the inductive employment of words meaning 'green', 'blue', etc., is to be preferred over those meaning 'grue', 'bleen', etc. If, as Peacocke says, a thinker's capacity of concept possession depends on the ways in which he is capable of 'finding it natural to go on', then the question becomes: could there be any *evidence* that establishes, objectively, why one should find one concept more natural than another, or detect one regularity rather than another in nature?

Goodman's riddle is predicated on the fact that there seems to be nothing about our evidence that suggests that our 'all emeralds are green' hypothesis is better confirmed than its 'grue' cousin. By the same token, there seems to be no evidence for why the former should in some way be preferable to the latter—but this sort of conclusion is unacceptable. For inductively employing ('projecting') predicates such as 'grue' means anticipating something which, according to our standard colour concepts and words, would amount to a colour change (cf. Mulhall 1989; see also Section 1.3.1 *infra*). It would seem that a projector of 'grue' displays, at the very least, a tendency to find rather different things "naturally compelling". This tendency will manifest itself in her preference of certain predicates over others, as well as in her finding certain expressions natural and others artificial or gerrymandered, etc. Given that he believes that the initial choice of language is arbitrary, Goodman maintains, consequently, that language (conceived as a *practice* rather than an abstract object) is ultimately the source of a hypothesis' inductive validity, in a manner which would have found some measure of approval from Wittgenstein. Goodman writes:

> …*the roots of inductive validity are to be found in our use of language.* A valid prediction is, admittedly, one that is in *agreement* with past regularities in what has been observed; but the difficulty has always been to say what *constitutes* such an agreement. The suggestion I have been developing here is that such agreement with regularities in what has been observed is a function of our *linguistic practices*. Thus the line between valid and invalid predictions (or inductions or projections) is drawn upon the basis of how the world is and has been described and anticipated in words. (Goodman 1983, p. 120-21; my emphasis)

The degree to which Wittgenstein would have endorsed what Goodman says here depends, one should think, on whether Goodman's adopts a "thick" or a "thin" reading of the term 'linguistic practice.'[33] Wittgenstein emphasized the necessity of agreement in *judgements,* as well that of agreement in behaviour, for people to count as participating in the same language game, and hence as belonging to the same linguistic community. A green- and a grue-speaker need not necessarily diverge in behaviour, although they would presumably disagree in judgements—in other words, their difference might be just a counterfactual difference, not a factual one. Questions of Wittgenstein exegesis aside, the quote from Goodman is quite remarkable in the way it swims against the tide of realist intuitions about induction. Surely, claiming that our language practice is the source of inductive validity must be putting the cart before the horse, for whether an inductive inference is correct or incorrect must be in the first instance a matter of how the way things are with the world, rather than how they are described in words? Similarly, a realist would expect there to be an objective fact of the matter about the way in which an inductive prediction agrees with observed regularities—i.e. an objective fact of the matter about (correct) description. And does not my use of the predicate 'grue' in connection with emeralds entail that I have certain specific expectations about how a certain portion of reality will be like in the future, expectations that can be confirmed or disconfirmed independently of the kind of predicates I chose to discribe them? Does the 'grue'-paradox not uncover,

---

[33] Wittgenstein *Philosophical Investigations*, § 241. My thanks go to David Levy for having pointed this out to me.

first and foremost, a problem about formulating an appropriate set of constraints upon the way we form our beliefs about the world?

This set of worries on the part of the realist—the philosopher who thinks that there *is* an objective fact of the matter determining whether we ought to think that emeralds are green rather than grue—links up very tightly with the worries a realist might have about Kripke's paradox. For the latter, too, goes against our realist intuitions concerning the relationship between language and the world. Here is how the two can be compared: Goodman says that a valid prediction of future events is one that *agrees with* observed past regularities. But this regularity, or sequence of events, is a finite one, and an infinite number of grue-like predicates "agree" with it, in the sense of having extensions such that the regularity in question can be correctly said to satisfy a description employing the predicate. Goodman's worry is why we should find certain predicates more descriptively *adequate* with respect to that sequence, or more 'natural', 'compelling', 'salient', etc., than others. Notwithstanding differences in point of departure and presentation, this problem is also Kripke's—the "regularity" in Kripke's case being expressed in terms of mathematical functions rather than colour predicates. Just as Goodman, Kripke underlines the fact that there are infinitely many possible ways of interpreting the pattern we might discern when we consider the "evidence", or in his terminology, the *facts* (whatever type of fact, external or internal, scientific or otherwise) for someone else or ourselves meaning something determinate by a word or symbol. We can 'fit' infinitely many descriptions employing infinitely many different predicates to that pattern, and it thus seems that we can ascribe infinitely many different meanings. Here, too, the claim is that we lack an independent account of why we *should* find certain predicates, and hence certain regularities expressed by their means, more compelling and descriptively adequate than others.

Both Goodman and Kripke thus search for a normative source that could define correct and incorrect 'fit' of descriptions to patterns, or regularities.[34] Both give similar answers. Goodman, as we have seen, claims that what constitutes agreement between past regularities and predictive hypotheses is a function of language use. Kripke, on the other hand, holds that the proper solution of the paradox paradox is to appeal to *communal agreement*, that is an agreement codified in the linguistic prac-

[34] For a useful discussion of the issue of the 'reality' of patterns, see Dennett, D. C. (1991). "Real Patterns" *Journal of Philosophy* **88**(1): 27-51.

tices of a community of speakers.[35] As above, a realist will complain that claiming that communal agreement is the normative source of correctness for meaning ascriptions must be putting the cart before the horse, because whether a hypothesis about a speaker's meaning is correct or incorrect must be in the first instance a matter of how the way things are with the world—in particular, with the speaker—rather than with whether anyone agrees with him! Even whole communities can be wrong. Similarly, a realist would expect there to be an objective fact of the matter about the way in which a meaning ascription agrees with observed regularities.

To summarize, Goodman's and Kripke's paradox capitalise on the fact that events or sequences of events fall under infinitely many different but extensionally correct descriptions, and that our way of extrapolating from finite sequences of events to the infinite extensions of predicates (and/or their meaning) which we use to characterise these events, is but one among infinitely many possible ones.[36] Intentional actions, *qua* events, are no exception to this, and nothing about the vaunted "special access" that we have to our own intentional actions—and hence to our intention to mean something particular by a word—is of help when faced with the Kripkean Sceptic. It is this last point that has prompted some to straightforwardly *identify* the new riddle with Kripke's paradox:

> All the evidence for the generalization "Emeralds are green" is equally good evidence for the generalization "Emeralds are grue" where, like Kripke's "quad", "grue" is a gerrymandered predicate. … I see a precise analogy between the idea that there could be evidence to confirm the greenness (but not the grueness) of emeralds and evidence that events are intentional under the description "adding" (but not under the description "quadding"). I would even say that the sceptical argument simply *is* Goodman's riddle of induction, tailored to field linguistics rather than mineralogy. (Allen (1989), p. 262.)[37]

[35] Cf. Kripke *Wittgenstein on Rules and Private Language*, p. 86ff. A heated debate has been raging among Wittgenstein scholars in the 80s on the question whether for Wittgenstein a linguistic practice is something that implies an actual community of speakers, as Kripke seems to assume, or merely one individual behaving in some regular way. (This is the issue whether a *radical* kind of Robinson Crusoe, a human being who has never been in contact with other human beings, could be following rules). We shall forego the pleasure of entering this arena, however, and also not engage with the question which relevant sense of 'practice' Goodman may have had in mind.

[36] For a useful discussion of the ontology of patterns, see Dennett "Real Patterns".

[37] Barry Allen alludes to W.V.O. Quine's famous field linguist undertaking the arduous task of radical interpretation (cf. Quine *Word and Object*, Chapter 1).

1.3 Curves and Redescriptions

If there are infinitely many possible ways of 'fitting' words to a given regularity or pattern of events, and no fact of the matter about the world allows us to decide between them, then the fit is *underdetermined* by the facts. Now, choosing one predicate rather than another in a description of a pattern or finite sequence of events is rather like the plotting of an infinite line over a finite sample of points—and it is precisely from this angle that the two paradoxes look like just one pair of shoes.[38] As a result of assimilating both paradoxes to a problem about curve-fitting, both philosophical problems are seen to ask the same question, What reason, if any, do we have to prefer "straight" predicates, or straight lines, over their gerrymandered cousins?, and thus to exploit the same type of underdetermination. Both simplicity- and similarity considerations turn out to play a role in the way in which we deal with this underdetermination.

1.3.1 The Curve-Fitting Problem

Our curve-fitting problem in the case of green vs. grue is the following one: our past and present visual observations of the colour of emeralds manifest a certain pattern, or regularity, namely they have all been green. Saying, on the basis of past observations of emeralds, that 'all emeralds are green' amounts to projecting that regularity into the future—or, graphically, to plot a line representing that generalization over the data points representing our observations to date of the colour of emeralds.

[38] For discussions of Goodman's riddle in terms of curve-fitting, see Goodman, N. (1970). "Seven Strictures on Similarity" *Experience and Theory*. J. W. Swanson, Boston, University of Massachusetts Press; Harman, G. (1994). "Simplicity as a Pragmatic Criterion for Deciding what Hypotheses to Take Seriously" *Grue! The New Riddle of Induction*. D. F. Stalker, La Salle, Illinois, Open Court**:** 153–71; Rubinstein, A. (1998). "Induction, Grue Emeralds and Lady Macbeth's Fallacy" *Philosophical Quarterly* **48**(190): 37-49; Forster, M. R. (1999). "Model Selection in Science: The Problem of Language Variance" *British Journal for the Philosophy of Science* **50**(1): 83-102. Although less common, Kripke's paradox is also sometimes discussed in terms of curve-fitting; see Anscombe, E. (1985). "Wittgenstein on Rules and Private Language" *Ethics* **95**: 342-52.

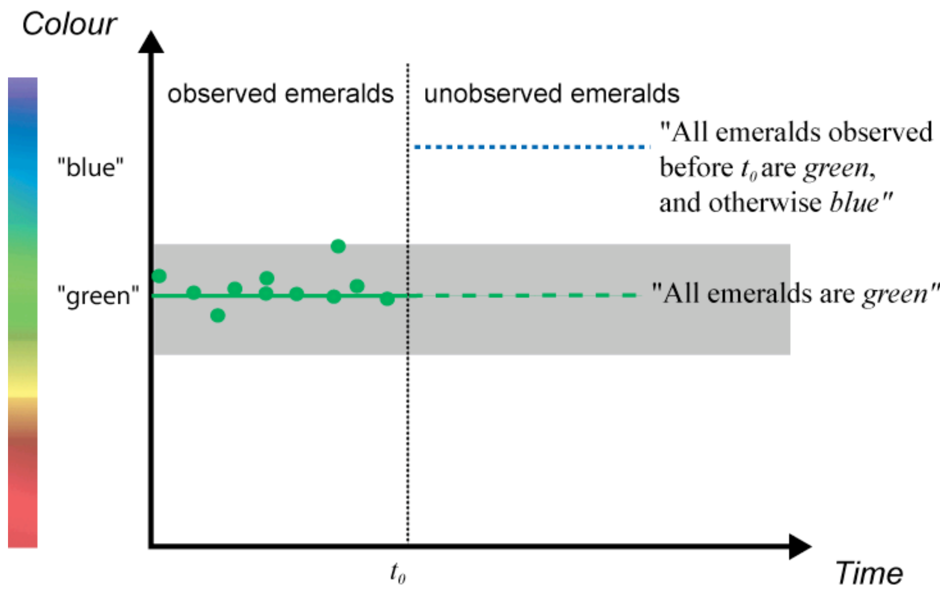

Fig. 1

The choice between the general propositions ‘All emeralds are green’ and ‘All emeralds grue’ thus amounts to a choice between two alternative continuations beyond t_0 of the “curve”—in this instance, a straight line—plotted over the data points. The grey rectangle represents the amount of chromatic variation conventionally allowed by the vague predicate ‘green’. (Given that we do not have a predicate for each single wavelength in the spectrum, ‘green’ in ordinary language does not apply to, say, 540nm light only; moreover, the boundaries of what is considered green are fuzzy, for there is no sharp demarcation in language between clear-cut instances of ‘green’, “borderline” cases, and clear-cut instances of ¬‘green’. All that matters, for present purposes, is that there *are* these clear-cut instances of ‘focal’ colours in which application of the predicate is more or less uncontroversial.[39] We’ll assume that observations or measurements of emeralds falling within the range indicated by the grey rectangle are *de facto* classified as uncontroversial instances of “All emeralds are green”).

[39] The problem of the vagueness of natural language is not the present one: ‘grue’ inherits its vagueness from its constituent terms, and there is no reason to suppose that a philosophical solution of the Sorites-paradox, if it was to be found one day, would not also take care of vague disjunctive paradoxes. The converse is not true: a solution of Goodman’s paradox would very likely leave the problem of vagueness as it is. For the notion of ‘focal colour’, see Rosch, E. and C. Mervis (1975). “Family Resemblances: Studies in the Internal Structure of Categories” *Cognitive Psychology* **7** and Sec. 1.3.3 *infra*. Finally, I should mention that natural light never comes in waves of exclusively one wavelength, but that it always forms a rich bouquet of unpolarized light of many wavelengths the exact composition of which influences the hue and subjective experience of colour. But these details are irrelevant to our (philosophical) discussion.

Conceptualizing Goodman's riddle as a curve-fitting problem, then, is interpreting it as the problem of how to map the independent variable 'time' onto the dependent variable 'colour'—or, if you prefer, onto the set of colour-predicates conventionally associated with a certain portion of the visible spectrum. What Goodman has shown is that although the choice of curve, given the data, is more or less *obvious* if we use 'blue' and 'green' on our colour-axis, it is *equally obvious* if 'bleen' and 'grue' are our primitives:

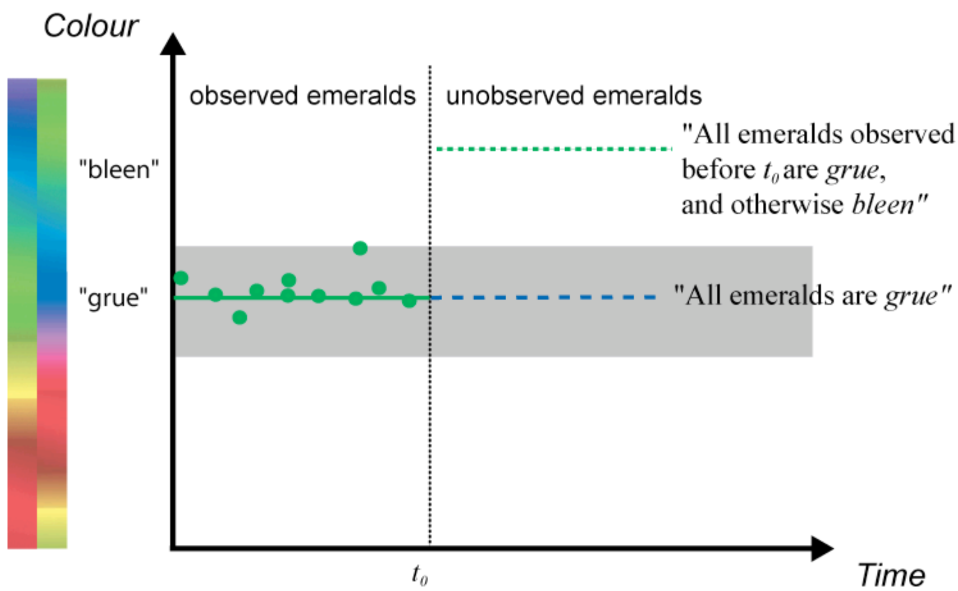

Fig. 2

It follows that we cannot determine which hypothesis is simple and which is complex—or which curve is continuous and which is discontinuous—just by *looking* at the relevant curves, for how particular curves will appear to us depends on how they have been represented. In fact, a suitably "gerrymandered" coordinate scale will allow us to represent any arbitrarily complex curve as a straight line, and any straight line as an arbitrarily complex curve. As Gilbert Harman puts it, you can represent *any* hypothesis as absolutely simple, for example by calling it 'H' (Harman 1994, p. 158).[40] It would seem that *which* system of representation we use, and conse-

[40] The attentive reader will have noticed something strange about my representation of the colour-spectrum in Figure 2. Whereas the left half shows, just like in Figure 1, the visible range of the electromagnetic spectrum from deep red (≈ 700 nm) to violet (≈ 400 nm), the right half displays an anomaly. It departs from yellow (≈ 565 nm–590 nm) and increases to the deepest shade of red (≈ 740 nm), then discontinuosly jumps to the other end of the visible spectrum, deep violet (≈ 380 nm), increasing again to the deepest shades of green (≈ 565 nm). (The borderlines between any of the colours are of course vague). What is going on here is that I have attempted to graphically, and inadequately, display what grue-type colours are like. The coulour coordinate scale is obviously gerrymandered, however, it

quently what will appear simple vs. complex, depends at least in part on what we are interested in. Harman notes that '"All emeralds are green" counts as simpler if one tends to be interested in whether emeralds are green. On the other hand, "All emeralds are green if first observed before [t_0] and are blue if not first observed before [t_0]" counts as simpler if one tends instead to be interested in whether particular emeralds are grue ..." (Harman 1994, p. 160).

The point of representing the 'grue'-hypothesis on a gerrymandered coordinate scale as in Figure 2 is to show that someone who makes inductive use of that predicate is just as *rational* as we are, at least insofar as her curve-fitting procedures are exactly same way as ours, with a marked preference for simplicity, i.e. straight lines. What makes her different from us is just that her concept of the colour grue—namely of an object being green until t_0, and afterwards blue—is just as simple for her as for us the concept of an object having a colour omnitemporally. This sort of difference is *not* to be confused with the difference between us and someone who looks at a sample such as

Fig. 3

and sees only *one* colour (i.e. someone who sees no perceptible difference between the left side and the right side). For, a grue-speaker would look at this figure and say, just as we would, that it is composed of *two* coloured rectangles—although for her, the colours are of course bleen and grue.

Goodman suggests that what will seem to us simple *vs.* complex, regular *vs.* irregular, or even similar vs. dissimilar, is consequent upon our choice of system of representation (i.e. our language), rather than the other way around.[41] If this thesis of

is difficult to properly represent in just which way: the point is that these are not actually colours, but "schmolours" (see *infra* on the concept of 'schmolour'), and it is strictly *impossible* to represent schmolours using colours.

[41] Cf. Priest, G. (1976). "Discussion: Gruesome Simplicity" *Philosophy of Science*: 43 432-437. The idea is of course by no means due to Goodman, and has been voiced by many philosophers before him, contemporary and less so. Some also ascribe it to the later Wittgenstein (cf. Wright, C. (1980). *Wittgenstein on the Foundations of Mathematics*, Cambridge, Harvard University Press, pp. 36-37; also *infra*). In his discussion, Goodman goes on to suggest that our choice of system of representation is, at least in the case of our inductive jugements, informed by pragmatic factors (Goodman *Fact, Fiction, and Forecast*, Ch. 4). Whatever the *truth* of these claims, the language-relativity of simplicity is clearly also an important assumption of Kripkean scepticism, and it is this parallel that we are presently interested in—*not* a defense of Goodman's position. The problem of the apparent relativity of simplicity

language priority is correct, and degree of complexity is language dependent, then, one should think, 'grue' should be ostensively definable. After all, it *is* perfectly simple for someone speaking an appropriate grue-language, and it is just like its cousin 'green', an observational predicate. This has been contested with the argument that in ostensive definition no reference could possibly be made to future change, chromatic or otherwise, of the sample, and that, therefore, the predicate could not be learned without one having first acquired 'green' and 'blue', plus an understanding of time (cf. Barker and Achinstein 1960). Indeed, this sort of difference looks like a plausible candidate for the fundamental difference between 'green' and 'grue', and the corresponding generalizations about the chromatic properties of emeralds. I have captured this idea in the graphs through the use of colours: in Figure 1 the colour of the continuous curve remains constant through time, in Figure 2 it *changes* at t_0. Surely, it is impossible to convey this sort of fact about grue things using purely *ostensive* means! Similarly, it should be impossible to convey the idea through ostensive definition alone that our sample in Figure 3 above is coloured in such a way that, after t_0, it will be *indistinguishable* from its colour-reversed counterpart—in other words, that the bi-coloured rectangle on the left of the dotted line in Figure 4 will be indistinguishable after t_0 from the bi-coloured rectangle on the right:

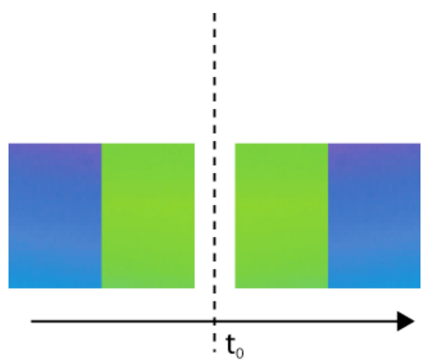

Fig. 4

A grue-speaker does believe that the rectangle on the left consists of a bleen and a grue rectangle, which should lead him to expect that after t_0, it would look exactly like the rectangle on the right. Surely, it must be impossible to *ostensively* convey this sort of expectation!

However, arguing in such a way is simply stomping one's foot and insisting on the impossibility of grue as a primitive, and thus ostensible, colour concept. A

due to choice of descriptive vocabulary continues to play a role in contemporary discussions of curve-fitting. Cf. Section 3.3.4.

grue-speaker, i.e. a person to whom 'grue' appears simple and 'green' complex, and who uses 'grue' inductively, is a truly exotic kind of cognitive agent, for she would also be a speaker who could *ostensively* learn the meaning of 'grue'. For there is no reason why a psychological agent—if she can acquire any concepts at all—should not be able to ostensively acquire a perfectly *simple* colour-concept. It is important, in this context, to not forget that to the grue-speaker, no colour-*reversal* takes place at t_0. Or rather, *her* concept of colour is such that no colour-reversal takes place: the left half of the rectangle is bleen, and the right half is grue, and this state of affairs remains unchanged as t_0 passes. What we would need, then, is not foot-stomping but a proof that grue as a simple colour-concept is impossible, and that a grue-language is somehow incoherent. This has been tried, not however with a degree of success that could be called unanimous.[42]

Kripke, in any event, fields an objection to the argument that we can ostensively define 'green' but not 'grue'using an appropriate green sample, which presupposes that a language that is entirely grue-like is thinkable. According to Kripke, the Sceptic can confute our efforts by further "gruefying" our language: 'It is no help to suppose that in the past I stipulated that 'green' was to apply to all and only those things 'of the same color' as the sample. The sceptic can reinterpret 'same color' as same *schmolor*, where things have the same schmolor if ...' [they have colour *x* before *t*, and colour *y* after *t*]' (Kripke 1982, p. 20).[43] Inevitably, a speaker who uses a language containing 'grue' as a primitive would also conceive of 'schmolour' as simple, and the *grue*-curve in Figure 2 would appear to such a speaker both continuous and of the same *schmolour.* If we grant this, then it is no help to try to distinguish colour from schmolour by holding, say, that things of the same colour, but not schmolour, are visually indistinguishable form each other. As has been pointed out by many authors, a Kripkean Sceptic can just continue to play his game *ad nauseam*, and for instance point out that for the grue-speaker, schmolours are *ingrustinguishable* from each other, where two things are ingrustinguishable from each other if they are indistinguishable before t, and distinguishable after, etc.[44]

---

[42] Serious doubts have been expressed over whether a grue-language is internally consistent or even conceivable, by Hesse, M. (1969). "Ramifications of 'Grue'" *British Journal for the Philosophy of Science* **20**: 13-25n; Shoemaker "On Projecting the Unprojectible"; and Mulhall "No Smoke without Fire: The Meaning of Grue".

[43] Kripke here echoes a claim by Ullian, J. S. (1961). "More on "Grue" and Grue" *Philosophical Review* **70**: 386-389.

[44] Shoemaker, S. (1975). "On Projecting the Unprojectible"Ibid. **84**: 178-219 argues that on pain of inconsistency, a putative grue-speaker would in fact need to possess a language where this sort of game has been played to its very end, i.e. until his entire language has been gruefied, along with its

It cannot be denied that for any claim of the form 'A resembles B with respect to C', the Sceptic can replace C (whether that be colour, shape, size, smell, microscopic structure, etc.) with an appropriate disjunctive term referring to some bizarre property that is shared by both A and B—hence apparently showing that anything resembles anything else to the same degree. Properties are plentiful, and by the same

---

concomitant conceptual scheme. Even on the assumption that this is possible (see Hesse "Ramifications of 'Grue'" for an argument that, at least as far as the language of science is concerned, it is not), it would lead to the consequence that the green-speaker and the grue-speaker would not be able to agree (after t_0) concerning the question whose induction had been falsified by the observation of an emerald after t_0. But then the grue-speaker would be deploying a grue-concept different from the one originally intended in Goodman's paradox, says Shoemaker, for the grue-concept in Goodman's paradox is intended to lead to clearly incompatible hypotheses. In other words, there is an 'agreement-after-t_0'-condition that would not be satisfied. It seems to me that this sort of proposal for a solution of the paradox is quite paradigmatic in so far as, like many others, it underestimates the absolute symmetry between the relevant sets of predicates, languages, or conceptual schemes. It therefore overlooks the fact that whichever philosophical argument against the paradox is deemed, *within* language L_1, to conclusively establish that the relevant rival set of predicates cannot be admitted into that language or its inductive practices, works equally beautifully from *without* it (namely within language L_2), conclusively establishing that the relevant set of predicates of language L_1 cannot be admitted into L_2 or its inductive practices. Thus, the grue-speaker can argue, à la Shoemaker, that given the way in which this funny predicate, "green", has been defined in his language (namely 'grue and observed before t_0 or not so observed and bleen'), a "green"-speaker, if he is really conceivable, would on pain of inconsistency need to "green"-ify his whole language (which is probably impossible for the language of science). But this would then lead to a lack of agreement, etc. etc. ... Another example for an author who underestimates in a similar way the symmetry between the languages is Jackson, F. (1975). "Grue" *Journal of Philosophy* **72**: 113-131. Jackson comes up with a 'counterfactual condition' on any colour-predicate—which has to do with the fact that it is part of our understanding of any colour-predicate that its satisfaction by an object does not depend on it being observed—that 'green' allegedly satsifies, but 'grue' fails. Jackson fails to see that if his counterfactual condition is stated in grue-language, then 'grue' satisfies it, whereas 'green' does not.

It is not our intention to defend the validity of Goodman's paradox, or of Kripke's for that matter, in this study, which is why these remarks have found their way into this footnote. However, it is worthwhile, for a proper understanding of Goodman, to understand the nature of the sort of language relativism that Goodman believed to have discovered with his riddle. It shows up in the way much of his later philosophy developed (cf. Goodman, N. (1976). *Languages of Art. An Approach to a Theory of Symbols*, Indianapolis, Hackett and Goodman, N. (1978). *Ways of Worldmaking*, Hassocks, Harvester Pr). In fact, taking into account his later writings, it is quite safe to assume, *contra* Shoemaker, that Goodman would have been entirely comfortable with the suggestion that green- and grue-speakers live in different *worlds*—or, as he would have put it, that they have constructed different *versions* of the one world—, versions that are both equally compatible with their respective experience. It would be naive to think that a simple test like waiting until the future arrives, and checking the colour of emeralds then, could suffice to extract a grue-speaker from his world and bring him "home" to ours. This is not what a "world" in Goodman's sense is about, for that world includes as an important part inductive practices which are found, from *within* that world, to be perfectly reliable. Thus, a grue-speaker will pick up the emerald, see that it is grue, that it has the same schmolour as previous emeralds, and that they are ingrustinguishable from each other—and he will go about his business as usual. This is precisely what Figures 1 and 2 were intended to illustrate: grue-speakers are, from an inductive point of view, people just like you and me: they like simplicity and straight lines, and nothing about future emeralds will ever make them think that they have been wrong about anything. Yet, they *are* wrong, for as every child knows, emeralds are not either observed before t_0 and green, or not observed before t_0 and blue. With every failed attempt at refuting the paradox, it becomes clearer and clearer just how difficult it can sometimes be to prove children right.

token so is similarity based on the sharing of properties.[45] Similarity with respect to logical or mathematical form is not in any way special: for every respect in which the use of '+' in the statement '2+2=4' resembles the use of '+' in the statement '100+2=102' (namely both being instances of adding), there is a parallel respect in which the way of using '+' exemplified in '2+2=4' resembles the use of '+' in '100+2=104' (instances of a suitably defined 'quadding'). The sceptic's challenge is to say what factual basis we have for choosing one such class of respects of similarity over any other one when we acquire the meaning of '+', and ascribe it to others and ourselves.

In fact, Kripke's quus-example, we may presume, is intended to hark back to a passage in Wittgenstein 1967's, where he describes an imaginary pupil, who has been given a series of examples and explanations intended to teach him how to add, and who is then instructed to go on and perform a series of additions of the form x+2. The pupil computes 2+2=4, 4+2=6, 6+2=8, etc., until he reaches 1000, and then, bizarrely, proceeds to compute 1000+2=1004. When questioned it turns out, in Wittgenstein's "thought experiment", that the pupil has not made a mistake. Rather, he has made the (to us) surprising decision that all the examples and explanations which he has been given up to then for how to correctly follow the instruction "Add 2!"—his *data* for our purposes—fit the simple function f(x)=x+2 less well than they do

$$\mathrm{g(x)} = \{\frac{x+2 \textit{ for all } x < 1000}{x+4 \textit{ for all } x \geq 1000}\}$$

… a curve-fitting exercise which we are bound to find irregular, but which he, bizarrely, finds quite trivial:

[45] As Lewis, D. (1983). "New Work for a Theory of Universals" *Australasian Journal of Philosophy* **61**: 343-377, reminds us, any two things share infinitely many properties, and fail to share infinitely many others; properties therefore do not capture facts of similarity. Unless they are what Lewis calls 'natural properties'; however, we also tend to believe that it is precisely the 'natural' properties that are inductively stable, which would threaten to render circular the use of similarity considerations in conjunction with the problem of induction.

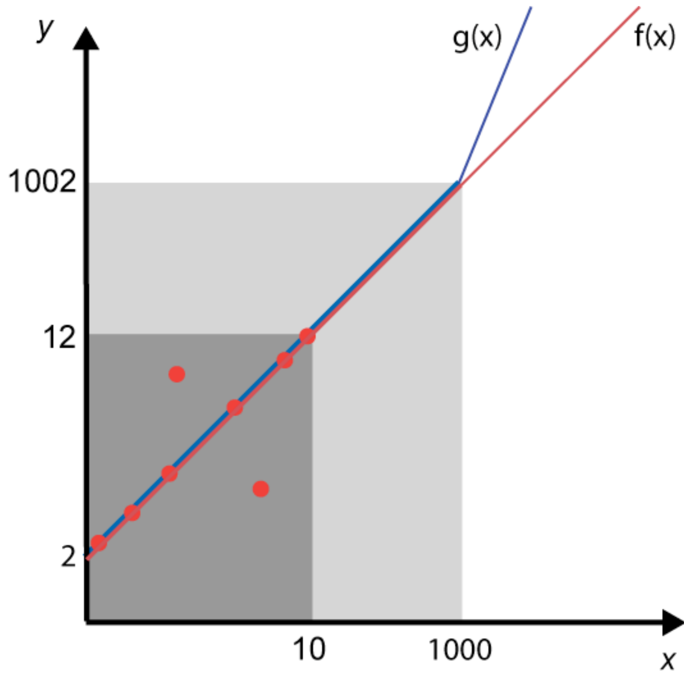

Fig. 5

g(x), to him, just seems *simpler* than f(x). Granted, the “data” in the pupil’s curve-fitting problem is likely to be of a more heterogeneous kind, comprising the teacher’s verbal explanations, examples, and other sorts of observations—but as Kripke’s *schmolor* example shows, we can think of the pupil as interpreting *any* given explanation, verbal or otherwise, as well as any concomitant observation, etc., so as to be compatible with the “bent” line. The dots that are not on the curve for either f(x) or g(x) are “outliers”, representing “noise” in the pupil’s data-set, for instance explanations that have not been understood, observations improperly interpreted, equivocal instructions on the part of the teacher, etc. Just like in any other curve-fitting exercise, the pupil decides to discard them when he makes his decision between f(x) and g(x).

The upshot of the curve-fitting analogy, then, is that the paradoxes employ the same *tool* for different kinds of effect. In both paradoxes, the occurrence of “gerrymandered” expressions is essential—both paradoxes ask the question what reason, if any, we have to prefer “straight” predicates, or straight lines, over their gerrymandered cousins, and thus they exploit the same type of underdetermination.

1.3.2 Multiple Redescription

What sort of underdetermination are we dealing with? As we have pointed out, any given object satisfies infinitely many different descriptions containing non-synonymous, though partially co-referential expressions. (A philosopher like Goodman would of course object to characterising multiple redescription in terms of non-*synonymy*, for synonymy is a relation between predicates involving something *more* than just the notion of their applicability to the same sort of things—it goes beyond extension. To Goodman redescription just *is* the correct application of an alternative predicate to an object.) The point of allowing for redescription in one's language is that of having a device for highlighting alternative features of the same object. Redescription is a way of differentially expressing one's various bits of knowledge about that object. This man over there can be differentially and correctly described as 'my neighbour', 'Mary Jones' husband', 'the president of the local ornithographical society', 'Mrs. Smith's lover', etc. Now, he can, of course, also be correctly described as 'a member of the species *Homo Sapiens* OR a member of the species *Homo Australopithecus*, 'an accountant AND not a prime number'—or, closer to our present problem, as 'a man with grue eyes'. We want to be able to say that there is something quite peculiar about the latter sort of descriptions.

The problem of how to define adequate constraints on redescription is not a specious problem confined to arcane areas of abstract metaphysics—it is central to any enquiry concerning the relation between language and the world. It also part and parcel of the philosophy of mind and cognitive psychology. Thus, one of the concerns of cognitive scientists is an account of how real-life knowledge comes to be so easily manipulable, flexible, and transferable by cognitive agents to other tasks (and other cognitive agents). According to one theory, the Representational Redescription Hypothesis (cf. Karmiloff-Smith 1992), initially implicit knowledge acquired through basic learning mechanisms is rendered progressively more explicit and abstract through a process of reiterated redescription, resulting in a hierarchy of increasingly explicit and more widely deployable representations. The challenge, for cognitive scientists, is to devise neural network models that perform redescription in the *appropriate* way, i.e. to find constraints on the process that mirror those hypothesized to be actually at work in human agents engaged in specific cognitive tasks. Our philosophical challenge, on the other hand, is to discover the nature of the constraints necessary for excluding the sort of sceptical reinterpretations we are concerned with. The two tasks are of course related.

Some Wittgenstein commentators, for example Anscombe 1985, pp. 345-47, think that—questions regarding the adequacy of Kripke's interpretation put to a side—Wittgenstein's rule-following considerations can be understood as precisely about the problem of (re)description, or rather about the problem of how something can still be considered the *same* thing in the face of its multiple redescribability. Given that we can describe what we do in infinitely many ways, even to *ourselves*, what makes it the case that two intentional actions constitute successive applications of "the same" rule, i.e. that they constitute "going on the same way"? Wittgenstein, in his rhapsodic style, frequently points to the connection between rule-following, description, and what constitutes sameness: 'Following according to the rule is FUNDAMENTAL to our language-game. It characterizes what we call *description*'; 'Disputes do not break out (among mathematicians, say) over the question whether a rule has been obeyed or not. People don't come to blows over it, for example. That is part of the framework on which the working of our language is based (for example, in giving *descriptions*)' (Wittgenstein 1978, VI, §28 and VI, §28, respectively; my emphasis); 'The use of the word "rule" and the use of the word "same" are interwoven' (Wittgenstein 1967, §225). Other Wittgenstein commentators emphasize that the rule-following problem is intimately connected with the problem of understanding our ability to *reidentify* objects—an ability dependent upon our ability to recognize the similarity of an object at t_1 to itself at t_0. Thus, discussing Wittgenstein's evolution from his earlier to his later "period", David Pears writes: 'In the *Tractatus* the reidentification of objects was never discussed, and all that was said about the way in which their names remain attached to them was that their names must respect their inherent possibilities of combination with other objects. But what about us? How do we recognize an object when we encounter it again? These questions led straight into [Wittgenstein's] later investigation of following a rule…' (Pears 1988, p. 148).

The concepts of description, similarity, and sameness are obviously closely interwoven. Two objects are similar if they are, at least in some respects, the same—if they are the same in all respects, then they are identical, i.e. one and the same object. Analogously for description and similarity: many philosophers allow the view that all circumstances correctly described as *F,* where *F* is any predicate, must be *similar*. If two circumstances are such that exactly the same set of predicates—or descriptions containing these predicates—is true of them, then they are the same circumstance. What separates Wittgenstein from other philosophers is that, like Goodman, he seems to believe that my judgements of similarity are *consequent upon* my judgements about the applicability of *F*, or the correctness of a description containing F, rather than the *basis* of such judgements. Crispin Wright explains:

> (…) our successive applications of an expression all seem quite familiar to us, at any rate when made sincerely. But the feeling of familiarity and the disposition to re-apply the same expression are all the same thing, so to speak. (…) whenever in the future I am prepared to call something 'green', I shall count it just on that account as being that kind of thing. Naturally, when I use the word 'green' sincerely on future occasions, it will seem to me that my use is familiar, is of the same pattern to which I earlier committed my-self; just that is incorporated in the adverb 'sincerely'. But how can I give sense to the idea that it *really is* of such a pattern, that there is an *objective similarity* in the circumstances in which I apply 'green' which God, for ex-ample, could discern? Wittgenstein wants to insist that actually we can give no sense to this idea. (Wright 1980, pp. 36-37; my emphasis)

Take Wittgenstein's example of the act of adding 2. In what sense is adding the number 2 to itself, i.e. to compute 2+2=4 *the same thing* as adding 2 to the number 1000, i.e. to compute 1000+2=1002? (Wittgenstein 1967, § 143ff.) What makes it the case that I have applied the same rule on both occasions, what (beyond the fact that we are disposed to describe the situation in these terms) makes it the case that I have "gone on in the same way" as before? N.B. the question regards our *act* of adding and the identity criteria for intentional actions, not the computations themselves. From the mathematical point of view, the equations '2+2=4' and '1000+2=1002' are, of course, related by the fact that both are instances of the application of the *plus* function to two different pairs of arguments. But this is a sense of 'application' in which there is no applier; the fact in question is a mathematical fact, namely that ⟨2, 2, 4⟩ and ⟨1000, 2, 1002⟩ are members of the infinite set of ordered triples constitutive of the *plus* function, and it obtains eternally without being "made true" by anyone.[46] Thus, we might want to say that what makes computing 2+2=4 and computing 2+1000=1002 the same sort of action is precisely the mathematical fact that ⟨2, 2, 4⟩ and are both members of the *plus* set. The trouble is, of course, that ⟨2, 2, 4⟩ and ⟨1000, 2, 1002⟩ are equally members of infinitely many *quus*-like sets, and the scepti-

[46] On the extensionalist construal of functionhood, *addition* simply is identical with that set (cf. Machover, M. (1996). *Set Theory, Logic, and their Limitations*, Cambridge, Cambirdge University Press). The view that mathematical facts are not "made true", but are true *simpliciter*, is philosophical majority vote at least since Descartes' *création divine des vérités mathématiques* found unfavourable reception (in other words, since philosophy turned rationalist).

cal question is precisely the question if there is anything about *me* and my intentions, or any other speaker and their intentions, that picks out one set rather than another, and justifies my claim that when I perform the above computations, I am computing the *plus* function.

A mathematical Platonist would express the relevant similarity between the two particular equations '2+2=4' and '1000+2=1002' in terms of their participating in the same mathematical universal. Not surprisingly, some philosophers who believe in universals also believe that our ability to apply a given predicate in a definite way is nothing but our ability to use universals. According to Simon Blackburn, 'Wittgenstein, Russell, and Goodman have simultaneously and independently developed modern ways of approaching the philosophical problems involved in understanding this kind of ability. Together they put immense pressure on our understanding of what it could be to assign a meaning of a predicate' (Blackburn 1984b, p. 69) Goodman's project, adds Blackburn, is the query of our right to take observed regularities in things as *representative*—in other words, it is precisely the query what authorizes us to think that we have got hold of an *objective* similarity in things when we apply the predicate 'green', a similarity with regard to which the future will resemble the past. Goodman's own answer to the question, involving as it does linguistic *convention*, presupposes that there is some fact making it true that we mean the one thing and not the other by our terms, says Blackburn (Ibid.). Blackburn thus thinks that meaning-scepticism is more fundamental than Goodman-style inductive scepticism (For Kripke's own assessment of that claim, see Sec. *infra*).

Wittgenstein preferred to approach the problem of accounting for what makes two different things things of the same kind by asking questions concerning the everyday use of the word 'same', questions like: 'How do I explain the meaning of "regular", "uniform", "same" to anyone?—I shall explain these words to someone who, say, only speaks French by means of the corresponding French words. But if a person has not yet got the concepts, I shall teach him to use the words by means of examples and by practice.—*And when I do this I do not communicate less to him than I know myself.*' (Wittgenstein 1967, § 208, my emphasis). Interpretation of these sorts of remarks is not exclusively a matter of Wittgenstein exegesis, for, Kripke himself points out, the comments on sameness and "going on the same way" play an important role in the internal economy of the Sceptical paradox:

> … when a teacher introduces such a word as 'plus' to the learner, if he does not reduce it to more 'basic', previously learned concepts, he introduces it by

> a finite number of examples, plus the instructions: "Go on the same way!" The last clause may … be regarded as vague, in the ordinary sense, though our grasp of the most precise concept depends on it. This type of vagueness is intimately connected with Wittgenstein's paradox' (Kripke 1982, p. 82, footnote).

There are, as Goodman says, 'countless lines of similarity' between two events that we can pick up on. Multiple redescription consists in choosing alternative predicates to make a given entitiy or series of entities appear similar—or the same, with respect to its satisfying that predicate—to other entities. This is how the use of '+' in 2+2=4 can be "the same as" the use of '+' in 1000+2=1004. In the course of his discussion, Kripke 1982 uses various kinds of redescription to illustrate the Sceptic's doubt. We have already encountered plus vs. quus, color vs. schmolour, chair vs. tabair. Sure enough, Kripke also employs 'grue', like this: 'Perhaps by 'green', in the past I meant *grue*, and the color image, which indeed was grue, was meant to direct me to apply the word 'green' to *grue* objects always. If the blue object before me now is grue, then it falls in the extension of 'green', as I meant it in the past' (Kripke 1982, p. 20).[47]

Just as in the case of the grue-paradox, it appears most natural to address this sort of silly philosopher's doubt by appeal to an *objective* notion of similarity, or sameness with respect to X. But there can be no answer to the sceptical challenge that relies explicitly or implicitly on the claim that an objective relation of 'sameness with respect to x' holds between the members of a given class of things, says Kripke, for this relation cannot be specified in ways invulnerable to the sceptic. Things, after all, can be the same with respect to their schmolor, schape, or schmathematical form, as it were, and the challenge levelled by *both* paradoxes is to say what factual basis we have for preferring one class of respects of similarity to any other one. In Goodman's case, for example, we would love to argue that 'green' is projectible whereas 'grue' is not, because we expect to be able to describe the resemblance between unexamined and examined emeralds in terms of the predicate 'green' rather than 'grue'—but this is precluded as long as we have not yet specified why 'grue' ought to be ruled out from our descriptive vocabulary. In other words, Goodman rejects in a manner exactly parallel to Kripke's the suggestion that in the emerald case, there is but one relevant type of similarity.

[47] Kripke, incidentally, adopts the Barker and Achinstein definition of 'grue' here, which is not Goodman's.

1.3.3 Similarity Relativized

Many philosophers nevertheless think that the intuitive and ultimately correct answer to Goodman's puzzle lies in the notion of similarity. W.V.O. Quine, for example, claims that 'Two green emeralds are more similar than two grue ones would be if only one were green. Green things, or at least green emeralds, are a kind. A projectible predicate is one that is true of all and only the things of a kind. What makes Goodman's example a puzzle, however, is the dubious scientific standing of a general notion of similarity, or of kind.' (Quine 1969, p. 116). Quine believes that the notions of 'similarity' or 'kind'—which he considers to be closely related—are too important to be left out of the scope of scientific inquiry. For, according to him, all general terms owe their generality to a resemblance among the things referred to. Thus, natural kind terms are best understood as referring to a class of objects that share certain important (objective) respects of similarity, and it behoves science to investigate these. Not only general terms, even the notions of 'cause' and 'disposition' may be defined in terms of similarity, or our "sorting of things into kinds", he says (Quine 1969, p. 116). The learning of language itself is conditioned upon similarity, he claims, because the correct use of a given word depends on our recognition of the resemblance between past circumstances in which the word was used and the present conditions in which it is to be applied (Quine 1969, pp. 116-117). Thus, the *ostensive* definition of, say, the word 'yellow' '… is a curiously comfortable case of induction, Quine says, a game of chance with loaded dice. At any rate this is so if, as seems plausible, each man's spacing of qualities is enough like his neighbour's. For the learner is generalizing on his yellow samples by similarity considerations, and his neighbors have themselves acquired the use of the word "yellow", in their day, by the same similarity considerations.' (Quine 1969, p. 125). Ostensive definition, and hence language acquisition, relies on implicit assumptions concerning the uniformity of the teacher's and learner's 'similarity spaces', concludes Quine.

Of course, the idea that language acquisition in general is a straightforward game of induction—albeit one in which the dice have been loaded—whereby the learner seizes upon regularities and similarities in the linguistic data available to her, has come under fire. Most generative linguists today think that Noam Chomsky's famous ‘Poverty of the Stimulus Argument’[48] has shown conclusively that language learning could not be just a special case of induction. The objection, in a nutshell, is that too many possible grammars are compatible with our primary linguistic data (the data which a language learner is typically provided with during childhood). The true grammar of a natural language, whether that be the grammar of English, Japanese, etc., invariably contains highly complicated and “unnatural” principles, principles which children would be very unlikely to hit upon on the basis of a standard inductive generalization (Laurence and Margolis 2001, p. 221). In particular, a child cannot simply opt for the most simple, or natural, set of principles, in the way we might expect it to when confronted with other tasks of inductive inference. “Simple” induction, in other words, cannot guide us when learning a language because the principles underlying language are too complex and “not natural”. (We might even be tempted to say that they are grue-like...). Moreover, the argument runs, we have empirical evidence for thinking that children typically avoid the sort of *mistakes* one would *expect* them to make if they simply generalized inductively upon the regularities and patterns they encounter.[49] The conclusion is that without being biased in a certain way—namely towards Universal Grammar, knowledge of which all humans are hypothesized to be genetically endowed with—a child would simply fail to become a competent speaker of its native language (Laurence and Margolis 2001, *ibid*). Given that the vast majority of children do reliably arrive at the right grammar and become competent speakers, they cannot therefore be unbiased, purely empirical inductivists, but rather must be assumed to have an innate endowment that restricts the space of possible hypotheses.

I take the question of whether children are little theoreticians with an innate bias towards Universal Grammar, rather than empirical inductivists without any significant innate linguistic knowledge—or indeed the question whether these two antagonistic accounts exhaust the space of possibilities—to be largely an empirical one. Perhaps the acquisition of new concepts is indeed impossible *tout court*, and absolutely all concepts are innate (as Jerry Fodor has famously and implausibly argued). In that case, learning a new concept is a process of fixing the parameters of a set of

[48] Cf. e.g. Chomsky *Knowledge of Language: Its Nature, Origin, and Use*. For a recent defense of the poverty of the stimulus argument against philosophical attacks, see Laurence, S. and E. Margolis (2001). “The Poverty of the Stimulus Argument” *British Journal for the Philosophy of Science*: 52(2) 217-276.

[49] Cf. the empirical examples described in Laurence and Margolis “The Poverty of the Stimulus Argument”, pp. 226-27 and 237. Of course, this sort of argument presumes that children fit straight non grue-like lines over their data.

hypotheses one has been endowed with from birth. These questions, I take it, are to be decided through scientific investigation (although there are good *prima facie* philosophical reasons to be sceptical).[50] Quine's claim that humans have an innate sense of similarity—although originally put forward in the context of his *inductive* conception of language acquisition—can be detached from this particular context. For the disagreement between nativists and empiricists does not concern the existence of innate endowments *simpliciter*, but rather their scope and domain-specificity (Laurence and Margolis 2001, p. 219). Thus, both die-hard empiricists as well as nativists will concur that:

> … it is likely that certain features of a situation in which a concept is applied will always strike a learner as dramatically more noticeable than others, so that certain possible hypotheses to account for the pattern of usage which he is experiencing will simply be ignored; and we learn a first language at sufficiently tender an age to make it plausible to suppose that our dispositions to be struck by certain features and to overlook others are largely *innate*, and so *largely* shared. (Wright 1980, p. 27)

Of course, Wright 1980 goes beyond this minimal agreement, and espouses a clearly inductivist interpretation of Wittgenstein's rule-following considerations and the problem of meaning acquisition (a position from which he has retreated in the meantime):

> The knowledge which we derive when we learn a first language is, plausibly, nothing other than inductively based conclusions about how expressions ought in general to be used, drawn from our experience of how they have been used. Thus to possess the same understanding of an expression as someone else will be to have formed, on the basis of suitable training, the same inductive hypothesis about its correct use.' (Wright 1980, p. 25).

Although this sort of view would make a rapprochement of Kripke's and Goodman's paradoxes trivial indeed, we shall not follow Wright's lead.[51] For whether or not word meaning is acquired inductively through a general all-purpose learning mechanism, or rather only with some substantial help from innate domain-specific endow-

---

[50] See e.g. Cowie, F. (1999). *What's within? Nativism reconsidered*, New York, Oxford University Press. A scientifically minded detractor might retort that there are good *prima facie* philosophical reasons to be sceptical of anything.

[51] Wright himself has considerably evolved away from his exegesis of Wittgenstein in the 80s (compare Wright (2002) *Rails to Infinity*).

ments, is, to reiterate, an empirical matter still under investigation and not for theoretical philosophy to stipulate. The centre of the present discussion is, rather, the brute fact with which all parties agree, that we *do* find some things more or less similar to each other than others. This in itself is fascinating enough, from a philosophical point of view. Our similarity space, whether innate or not, is part of what Wittgenstein used to refer to as 'bedrock'. At bedrock, there is a disposition to be struck by some features of the objects impinging on our sensory surfaces, and not by others—a disposition which is likely to be at work both in the inductive inferences we then proceed to make over these objects, as well in the process of meaning-acquisition, whether the latter is itself inductive or not. (How could the acquisition of syntactical structures work if we did not even pick up upon any similarities between two successive instances of the phonetical shape c^a^t ?). What we are concerned with here, then, are the philosophical implications of 'bedrock'.[52]

The (philosophical) importance of our sense of similarity surfaces in many adjacent fields of inquiry. Donald Davidson, for example, writing about the metaphoric use of words, points out that the characteristic of a metaphor is the fact that it '… makes us attend to some likeness, often a novel or surprising likeness, between two or more things. Ordinary similarity [in contrast] depends on groupings established by the ordinary meanings of words. Such similarity is natural and unsurprising to the extent that familiar ways of grouping objects are tied to usual meanings of words.' The (quasi-Whorfian) view[53] that ordinary similarity *depends* on classifications established by ordinary word meanings, i.e. on language, has been overtaken by advances in cognitive psychology: whether nativism is true or not, it is likely to be the other way around. However, the idea that what we are dealing with in the case of the "similarities" expressed in terms of 'grue' might be, in Davidson's terms, *unfamiliar ways of grouping objects tied to unusual meanings of* (unusual) *words*, is fruitful in our

[52] There is, incidentally, empirical evidence for the fact that what initially lies at 'bedrock' can subsequently be modified through culture. Studies have shown that although Japanese babies are born, like all other babies, with the innate ability to perceptually discriminate between phonemes ^l^ and ^r^, they subsequently loose this ability. The accepted explanation for this is that Japanese does not rely on the phonetic difference between the two, so that whether a baby hears ^l^ or ^r^ will never make a difference to the morpheme it hears, and hence never translate into a difference of meaning. Given that the ability is not used, the brain de-activates it.
Again, we are not concerned with exactly delineating the aetiology of 'bedrock' and the question of nativism vs. empiricism, as for instance in the question whether one could conceivably become a grue-speaker through training. Our interest lies, rather, in the fact that one is a grue-speaker because one's 'bedrock' differs (and not one's rationality). See *infra*.

[53] Cf. Whorff, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. Edited by John B. Carroll*, Cambridge, Massachusetts, MIT Press; and next footnote.

context. For if it is in fact the other way around, then unusual meanings of words will be acquired through one's tendency to group objects in unfamiliar ways, and we are pointed towards the idea that anyone who uses grue simply must have built his linguistic and conceptual house on an entirely different sort of foundation—his 'bedrock' is not ours. Thus, it seems plausible that a shared, and probably innate, endowment with a sense of similarity will see to it that most *humans*, at least, would find blue emeralds after *t* much less similar to emeralds examined before *t,* than green emeralds (as Quine pointed out). They would find them *so* different from green emeralds before *t* that they would be enormously surprised and conclude that this was one of the egregious cases in which induction had failed them; in brief, humans project 'green' and not 'grue'. It seems that it is precisely our sense of similarity that prohibits us from conceiving of 'grue' as a non-disjunctive colour predicate, i.e. a predicate that does not imply that the objects it applies to undergo a change of colour.

Anthropologists working with indigenous people from non-technological civilizations have shown that the sense of similarity we are interested in here has nothing to do with the way our particular colour words happen to carve up the light spectrum (cf. Rosch 1977). The same 'focal colours' are perceptually salient and easier to memorise for humans across cultural and linguistic differences—even, as Rosch showed, for the Dani in New Guinea, who dispose of only two colour terms. Even those humans who differentiate colours with terms roughly equivalent to, perhaps, our words 'dark' and 'light', will find certain "central", or focal, types of colour particularly easy to recognise in memory and naming tests. And they certainly are as good as we are at discriminating between equally bright patches of, say, orange and green. Nothing else should have been expected, of course. It is widely accepted today in cognitive psychology that a lot of our 'categorical perception' (among which perception of colours) is innate, and that it has been acquired through Darwinian evolution.[54] Whether we actually have *names* for our colour categories, however, is mani-

---

[54] Stephen Harnad explains: '... until the discovery of the physiological bases for color vision (Boynton 1979, DeValois & Devalois 1975) the only theory of color CP [the categorical perception of colour] was the Whorf Hypothesis (Whorf 1964), according to which the location of color boundaries is determined by where languages happen to put them. This was almost too nonspecific a hypothesis to generate focused research (and once it did, it turned out that color boundaries were largely determined by species-specific color receptors rather than by language' Harnad, S. (1987). "Categorical Perception: The Groundwork of Cognition" *Categorical Perception: The Groundwork of Cognition*. S. Harnad, New York, Cambridge University Press; accessed 03/05/2003 at http://www.ecs.soton.ac.uk /~harnad/Papers/Harnad/harnad87.cpreview.html. Even though the hypothesis that language determines how things look to us has been falsified (at least for colours), this does not mean, as Harnad points out, that (supervised) learning cannot lead to a change in our categorical perception, to a "warping" of our similarity spaces—an effect equally well documented. The verdict, therefore, seems to be

festly influenced by further factors. What matters for present purposes is that Goodman's paradox can easily be constructed for the Dani, because they, too, will find green objects before *t* more similar to green objects after *t* than to blue ones after *t*, irrespective of the fact that they have no terms for these colours. Presumably, they will project their two only colour predicates in the same, regular, way as we do. In other words, an individual from such a civilisation, even though possessing radically different colour concepts, is much more like us than any Goodman-type projector of 'grue', because (1) her sense of similarity, and (2) inductive practices, are like ours. She shares with us the same 'bedrock'. What distinguishes the possessor of the *grue*-concept from us is that he does not.

The significance of all this for our subject can be summed up as follows: the concept of 'similarity'—or, as Wittgenstein would have put it, of 'going on the same way'—provides a way of accounting for both the type of "projection" involved in inductive inferences that preoccupied Goodman, and the "projection" present in the acquisition and subsequent attribution to others and to myself, of meaning. From the perspective of the recognition of similarities, the process of choosing the predicate 'green' rather than the predicate 'grue' to refer to an inductively stable property of observed emeralds (call it 'projecting$_G$'), is the same as the process of abstracting the 'plus' rule rather than the 'quus' rule from the series of numbers '0, 2, 4, 6, 8, 10, 12…' and any accompanying explanations (= 'projecting$_K$'). Now it may very well be the case that we have an innate simple concept for *plus* and lack a corresponding one for *quus*, which we need to assemble from simple ones. In that case, Wittgenstein's pupil would not properly speaking *inductively* infer from, say, the series '0, 2, 4, 6, 8, 10, 12…' to the set of even numbers. Rather, this data would serve as the stimulus for the 'activation' within him of his innate concept for *plus*. Nevertheless, this process is a process whereby the pupil begins with a limited sample and ends up with a concept or a word meaning, and thereby—due to the infinitary nature of concepts and meanings— with something that uniquely determines an infinite sample. The suggestion mooted in this section was that any process which pairs up one-to-one a series of finite data with an infinite object, constitutes a projection that is still a form of *curve-fitting*.

mixed: yes, some parts of our sense of similarity are innate and species-specific (e.g. colours), but others are not (e.g. our Japanese babies in Footnote N°50; Harnad's example is that of the acquired capability of factory workers to distinguish male chicken from female ones).

Wittgenstein liked to emphasize that recognising a particular similarity in a given set of objects is the *essential ingredient* in learning how to follow a rule: 'For us a series has a face!', he claimed (Wittgenstein 1967, §228), and he seemed to mean it quite literally. The only difference between a plus-rule follower and a quus-rule follower is precisely that the latter has extracted a different type of similarity from the samples of words, symbols, pictures, or other kinds of experiences, that were presented to him during his period of apprenticeship. The sample "speaks" differently to him, as Wittgenstein put it—he also used the expression 'the rule's mouth'. Davidson, writing in the context of a discussion of Socrates' dialectic method of philosophising, describes the situation as follows:

> In learning a first language, many words must be learned by ostension … . Ostension has an obvious limitation: in our whole lives we can be exposed to no more than some finite number of examples. *There is always a chance that when a new case arises the learner will deviate from the norm.* ... in the learning situation the deviant learner is simply someone who, perhaps wisely, has been persuaded for the moment to suit his practice to that of one or more others. *A stubbornly deviant learner,* on the other hand*, may have an insight into a deep similarity of cases that others have missed*, and she may carry the community with her. This is exactly what Socrates does, or attempts to do, when he tries to persuade his companions to stop using the word "just" to apply to acts in which someone returns harm for harm, and to apply it instead to acts that return benefit for harm. (Davidson 1994, pp. 435-36; my emphasis)

The importance of the question of our "insight" into the similarity common to a series of cases—or of our disposition to be struck by some features and not by others—during ostension definition,[55] is what prompts Kripke to declare that 'Nevertheless, intuitively it does seem clear that 'grue' is positional in a sense that 'green' is not. Perhaps that sense can be brought out by the fact that 'green', but not 'grue', is learned (learnable?) ostensively by a sufficient number of samples, without reference to time.' (Kripke 1982, p. 59n). Kripke goes on to say that the proper defence of Goodman against this argument would be a Sceptical move as follows: 'Who is to say that it is not 'grue' that others (or even, myself in the past?) learned by such ostensive learning?' (Ibid.), and comments that this leads directly to the Sceptical problem

[55] Ostensive definition is a philosopher's of language favourite pet, because it is of course the only sort of definition that allows to introduce primitive terms without relying on antecedently understood terms, and thus, presumably, without falling prey to the sort of infinite-regress-of-rules argument we have outlined above.

about meaning. Hence 'serious consideration of Goodman's problem is impossible without consideration of Wittgenstein's'.

I agree, insofar as serious consideration of neither paradox is possible without reflection on the role of similarity, the steps and transitions that we find 'primitively compelling', or curve-fitting. Neither paradox is, from that perspective, more fundamental than the other, however, for both Goodman's projector of 'grue' and Kripke's 'quus'-speaker, are "deviant learners" in exactly the same sense, or bent rule-followers who have a differing insight into similarities.

The difference between the paradoxes is thus very shallow indeed: it is merely a difference in the particular *type* of rule at hand, or the kind of concept involved. This difference does not go to the core of the matter. Unfortunately, Carnap's reply to Goodman has initiated a tradition of reacting to the new riddle of induction which *overlooks* that it does not rely substantially on a specific kind of gerrymandered predicate, in particular on one with a built-in reference to a temporal ordering. Yet, already in 1946, Goodman anticipated this sort of response, and explained that the 'positionality' of grue is not an essential feature of the paradox—'grue' might just as well be defined as 'green, or under the Eiffel Tower and blue', or indeed *without reference to any ordering at all,* be it temporal, spatial, numerical, or otherwise (Goodman 1946, pp. 383-84). Barry Stroud gives a good example to illustrate how paradox obtains for just about any kind of disjunctive predicate:

> All the emeralds anyone has observed or ever will observe will be observed emeralds, so all the 'evidence' there will ever be from observation is compatible with the hypothesis that all emeralds are green-if-observed-and-otherwise-blue. And all the things that are ever referred to as 'otiose' will be things that are referred to, so there is nothing in the 'evidence' there will ever be from actual applications to favour the hypothesis that 'otiose' means without a function over its meaning without-a-function-if-referred-to-and-otherwise-hateful. (Stroud 2000, p. 129)

2. Realism about Dispositions

A common reply to Kripke's sceptical argument is the appeal to (facts about) dispositions. Surely, it is argued, what distinguishes a speaker who means *plus* by '+' from another who means *quus* ought to be what they are disposed to do with that symbol, in other words a difference in their respective dispositional states. Recourse to the notion of disposition is, as Kripke comments,

> … a response that I have heard more than once in conversation in this topic. According to this response, the fallacy in the argument that no fact about me constitutes my meaning plus lies in the assumption that such a fact must consist in an occurrent mental state. Indeed the sceptical argument shows that my entire occurrent past mental history might have been the same whether I mean plus or quus, but all this shows is that the fact that I meant plus (rather than quus) is to be analyzed dispositionally, rather than in terms of occurrent mental states. (Kripke 1982, p. 22)

It is precisely the non-occurrent, modal, nature of these mental states that distinguishes *plus*-speakers from *quus*-speakers, and makes facts about meaning dispositional facts. Thus, although none of my actual thoughts and other mental states allow me to differentiate between the plus and quus hypotheses, the idea is that '... there were dispositional facts about me that did make such a differentiation. To say that in fact I meant plus in the past is to say—as surely was the case!—that had I been queried about '68+57', I would have answered '125'. By hypothesis I was not in fact asked, but the disposition was present nevertheless.' (Kripke 1982, pp. 22-23).

This chapter will examine some important realist accounts of dispositions and scrutinize, in particular, the way in which they go about solving Kripke's challenge.[56] Kripke's paradox will serve as a touchstone for evaluating the advantages and disadvantages of realism about dispositions. The question we shall eventually put to each of the proposed solutions, or sometimes mere sketches of a solution, is how well they cope with our contention that Kripke's and Goodman's paradox are but two sides of the same coin. To appreciate the nature of the change brought about by the new real-

[56] This chapter makes no claim to be a complete survey of contemporary realism about dispositions. The discussed accounts have been chosen on the basis of whether their authors have explicitly addressed Kripke's paradox.

ism, however, we need to first take a look at the traditional framework provided by Carnap and Goodman, within which philosophical discussion of dispositions still takes place. This will also be the occasion to clear a few common misconceptions of the empiricist view of dispositions. Thus, Section 2.1.1 presents the traditional empiricist approach to non-occurrent properties in terms of conditionals, namely Carnap's famous 'reduction sentences'. I argue that Carnap does *not* hold the implausible view, often ascribed to him, that if a particular consists of a substance that has never been subjected to the test-condition for a disposition, then the very question whether the particular has that disposition is meaningless. Rather, Carnap thinks that disposition-ascriptions under these circumstances are, to some degree, conventional. In 2.1.2, I compare Goodman's approach to Carnap's, and show how Goodman turns the problem of permanently unmanifesting dispositions into the problem of confirming relevant law-like generalisations. Goodman's account suffers not from fallaciously trying to reduce dispositions to possible events, as frequently held, but rather from staking too much on the availability of (strict) law-like generalisations, as well as threatening to make dispositions causally inefficacious.

Section 2.2 then explores whether realist construals of dispositions can do any better, in particular with respect to our intuitions concerning permanently umanifesting dispositions. 2.2.1 expounds Martin and Heil's arguments in favour of realism about dispositions, and their recent dispositionalist solution to Kripke's paradox. 2.2.2 scrutinizes the substantial metaphysical commitments of Martin and Heil's theory, such as a dispositional state's intrinsic 'projectivity'. Section 2.2.3 then argues that the account on offer is insufficient against Kripke, for no indications have been provided how to satisfy the Sceptic's demand for factual evidence that a given agent has a *plus*- rather than a *quus*-disposition. Then, D.H. Mellor's conditional, but nevertheless realist, theory of dispositions is credited with essentially the same sort of shortcoming. Section 2.3.1 presents Mellor's case for maintaining a conditional analysis of disposition predicates in the face of Martin and Heil's counterexamples, and Section 2.3.2 takes a look at Mellor's Carnap-style reduction sentence solution of Kripke's paradox. We find that the fundamental idea in both examined realist approaches is the same: our dispositional states ensure, *in the absence of defeaters,* that whenever we are computing '*m*+*n*', we obtain the correct answer. Even though defeaters are always and necessarily present due to our finitude, this does not weaken the disposition's entitlement to be considered real and wholly present. Section 2.3.3 applies the argument previously wielded against Martin and Heil to Mellor: the difference between a *plus*-speaker and a *quus*-speaker is purported to consist in a specific conditional fact, described by a ‘reduction sentence’, assumed to be true of the former but not the latter. However, Mellor fails to provide instructions on how one could establish, whether in practice or in principle, that a conditional fact obtains in a given case, and his theory looks rather stipulative for it. The source of the difficulty lies in the circumstance that the required fact is a *theoretical* one.

Section 2.4 elaborates on this theme by examining the explicit (theoretical) presuppositions of Millikan’s teleological solution of Kripke’s paradox. Section 2.4.1 begins with an account of Millikan’s teleological theory of rule-following. Millikan theorizes that we have evolved a competence to follow the *plus*-rule, and to mean *plus* by ‘+’, rather than any *quus*-like rules. She acknowledges that this is, at least, the hypothesis that *best explains* our species’ survival. A central role in Millikan’s account is played by the distinction between actual performance and competence. The last Section of the chapter, 2.4.2, focuses on this distinction and on the process by which we infer competence from performance. I argue that there is a structural similarity in all realist accounts examined in so far as all are in fact built on a competence/performance distinction or its equivalent, and that the sort of inference that leads us from facts about the latter to the postulation of facts about the former, is the same throughout. I close the chapter with the observation that this inference might be identical with the sort of inference we make in the cases examined in Goodman’s paradox. The inferences in question contain substantial idealizations.

2.1 Empiricism about dispositions

The empiricist approach to dispositions and dispositionality departs from our naïve understanding of what they are, from what might be characterised as two pre-theoretical, and non-negotiable, facts about dispositions. The first is that dispositions are features of things generally inaccessible to direct observation. As Wilhelm Essler points out, I can see that the window over there is dirty, but not that it is fragile; I can find out through simple inspection that a man is tall, but not that he is courageous, etc. (Essler 1970). The not immediately observable and non-occurrent, features, aspects, properties, of a given object, its ‘solubility’, ‘courage’, ‘alcohol abuse,’ or ‘instability’ are usually labelled its dispositions. In fact, the word ‘disposition’ presents itself as a rather vague umbrella-term that embraces many things we have various other names for, such as ‘tendency’, ‘habit’, ‘capacity’, ‘liability’, etc., most of which receive different definitions from different authors, or remain ill-defined. However, the fact that a dispositional state is unavailabile to direct observation (in contradistinction with a disposition’s manifestation event), and the corresponding impossibility to ostensively define a disposition-predicate, represents a minimal consensus among all authors on the subject. A “categorical” (non-dispositional) property may be non-occurrent right now—the tide may be low now—but this does not mean that I cannot ostensively define ‘high tide’ by waiting until it occurs, or by somehow making it occur.

If a dispositional state is not observable, what makes us believe that it is a state, a something, in the first place? This is the second pre-theoretical “fact” about dispositions: most people would, if asked to explain what statements ascribing dispositions to things mean, reply by using some type of conditional construction. Thus, the meaning of ‘*a* is fragile’, for example, is likely to be explained as follows

> “a is fragile” means that “if a were stressed, a would break”

Whether correctly or not, dispositional properties are commonly understood as properties that manifest their presence in a particular way if certain conditions obtain. Of course, there are many cases of disposition ascription that are less straightforward. Take for instance the claim that *this* man sitting over there is courageous. Here, the ascription does not, at least not *prima facie*, amount to any precise claim of the form if X happens, then this man will do Y. However, an adequately informed speaker might nevertheless say something like “in situation X in the past, this man has done Y, and all people who do Y in situation X are (deserve to be called) courageous.” In other words, the speaker might assent to

> “*a* is courageous” means that “if *a* were in situation (of type) X, *a* would act in way Y”

Counterfactual conditional explanations of dispositional predicates are, in one form or another, endemic—they are how we explain their meaning to ourselves in everyday life. All theories of dispositions need to either take this fact on board by accepting a conditional analysis of disposition predicates, or somehow explain it away.

The empiricists felt of course that these features of dispositions and their ascription made them quite problematic and mysterious.[57] They saw both a disposition's strange intrinsically non-observable, "ethereal," character, as well as the widespread use of non truth-functional conditionals in its ascription to objects, as important obstacles on the way to a sober scientific world picture. Thus, empiricist philosophers usually stipulate that the question whether a given disposition is present or not, and hence whether a given dispositional predicate is correctly applied or not, should be considered tantamount to the question whether the object of ascription satisfies a given operational test. The general idea is something like

$$D(x) \equiv \forall(x)(C(x) \rightarrow M(x)),$$[58]

where M is a predicate characterising a type of event called D's standard "manifestation" or "confirmation" event, and C is D's "triggering" or "stimulus" condition. The operational test for x's possession of the disposition thus consists in making it the case that C(x), and to observe whether M(x) occurs. The immediate difficulty with any such account is evidently what to do if C(x) does not obtain. Both Carnap and Goodman have similar ways for dealing with cases where the test condition for a disposition does not, or cannot be realized. The shortcomings associated with their conditional account is what will prompt the realist "revolution" in the theory of dispositions.

2.1.1 Carnap

There are of course various ways in which an empiricist can add flesh to the bones of the above formula. The most influential one was Rudolf Carnap's (Carnap

---

[57] I follow Wright, A. (1991). "Dispositions, Anti-Realism and Empiricism" *Proceedings of the Aristotelian Society* **91**: 39-59, in using the term 'empiricist' in this context to denote all philosophers with an approach to metaphysical questions based on the doctrine that the only viable epistemic foundation of metaphysical claims are observable occurrences. What 'empiricism about dispositions' in particular amounts to shall be described in the following.

[58] Mumford, S. (1998). *Dispositions*, Oxford, Oxford University Press, p. 16, calls this the 'empiricist view of dispositions'. Given that non-empiricists are free to also endorse an account of dispositions in terms of conditional sentences, we might perhaps more aptly call it the *operationalist* approach, because it is meant to specify a (hopefully practically realisable) test condition, the obtaining of which allows to ascertain whether or not a given object has a given disposition. If we interpret '→' as material implication, the analysis amounts to a partial explicit definition of '*D*' (see *infra*).

1936-37).[59] Carnap felt that we need to avoid a non-truth-functional interpretation of the '→' operator, for only truth-functionality would allow him to define unique introduction rules for disposition-predicates into a properly formalised language of science, L. However, if we straightforwardly interpret '→' as material implication, we cannot explicitly define dispositional predicates. As Carnap points out, defining 'x is soluble' along these lines as meaning the same as 'whenever x is put in water, x dissolves,' or

$$S(x) \equiv (\forall t)(C(x,t) \supset M(x)),$$

(where C(x,t) represents the test condition 'x is put into water at t,' and M(x) the reaction 'x dissolves'), leads to undesirable results: it is obviously not the case that everything that is never put into water is soluble (cf. Carnap 1953, p. 53).[60] The same problem will reoccur in all cases in which a given disposition's test conditions never obtain, whether contingently or necessarily so.

Carnap's famous proposal to circumnavigate this difficulty is to renounce explicit definition and to use so-called 'reduction sentences' for introducing new (non-primitive) predicates into L. Every such predicate, e.g. S(x), is to be introduced (= its meaning is to be given) by means of two sentences, which specify experimental conditions that establish whether or not a given point in space and time exhibits the property described by S(x). For example, in order to find out whether S(x) = 'is soluble,' correctly applies to x at t, we need to ascertain the truth of two sentences of the form

$$C_1(x,t) \supset (M_1(x,t) \supset S(x))$$
$$C_2(x,t) \supset (M_2(x,t) \supset \neg S(x))]$$ (cf. Carnap 1953, p. 53)

where $C_1(x,t)$ and $C_2(x,t)$ are describing specific tests for S(x), such as 'x is put into water at t', and $M_1(x,t)$ and $M_2(x,t)$ are the possible results of the experiment (e.g. 'dissolves' and 'does not dissolve').[61] In the special case in which the experimental conditions necessary to establish that S(x) are the same as those for establishing that

[59] Page references will be to Carnap, R. (1953). "Testability and Meaning" *Readings in the Philosophy of Science*. H. Feigl and M. Brodbeck, New York, Appleton-Century-Crofts.

[60] Cf. also Mellor, D. H. (1974). "In Defense of Dispositions" *Philosophical Review* **83**: 157-181. This is sometimes referred to as "Carnap's Paradox." (e.g. Mumford *Dispositions*, pp. 46-50).

[61] Given that this introduces a method for finding out whether x is S by finding out whether x is C_1, M_1, C_2, and M_2, the method "reduces" 'S' to 'C_1, M_1, C_2, M_2', says Carnap (Carnap "Testability and Meaning", p. 53). The important *caveat* is that this does not work if C_1 and C_2 never obtain (see *infra*.)

$\neg$S(x), as with 'soluble' and many other disposition predicates, the reduction pair of sentences becomes equivalent to what Carnap calls the 'bilateral reduction sentence'

$$\forall(x)\forall(t)\ [C(x,t) \supset (S(x) \equiv M(x,t))],$$

which reads as 'if anything x is put into water at any time t, then, if x is soluble in water, x dissolves at the time t, and if x is not soluble in water, it does not' (Carnap 1953, p. 53). This amounts to only a *partial* explicit definition of S(x), for the sentences only explicitly describe those conditions under which the predicate does apply to a certain space-time point, but leave it open whether it applies under others. Elsewhere, Carnap refers to reduction sentences as *conditional definitions* (cf. Carnap 1956, p. 53):[62] in case neither $C_1(x,t)$ nor $C_2(x,t)$ ever obtain, the reduction sentence is not valid, and it remains *indeterminate* whether x is S, thus avoiding the implausible consequences mentioned above.

Notoriously, Carnap's proposal has not been well received. A contemporary author remarks that 'The formal reduction sentences … were thoroughly dissected by an impressive array of critics … Even the mention of Carnap's name brings to the minds of many (you can see the red lights flashing) all the shortcomings and problems associated with his construal of disposition predicates.' (Nordmann 1990, p. 381) It is neither within the scope of this thesis nor necessary for its purposes to examine all shortcomings associated with this account of dispositions. Instead, we shall focus our attention on a series of comments which Carnap himself makes about his proposal. Reduction sentence-pairs or bilateral sentences, not being explicit definitions, establish the meaning of the predicate introduced only for the cases in which the test condition is fulfilled (Carnap 1953, p. 56). He goes on to point out that we need of course an account of dispositions that allows for dispositions that *never* manifest themselves. For instance, we want to be able to say of a wooden match, which was lit at time *t* and subsequently burnt, that although it was never actually placed in water before *t*, the question of its solubility at that time is *not* indeterminate. Being made of wood, we know that it could not have been soluble. The indeterminateness on Carnap's account of dispositional predicates whose test conditions fail to obtain needs to be *reduced*. This can be done

[62] Quoted by Nordmann, A. (1990). "Persistent Propensities: Portrait of a Familiar Controversy" *Biology and Philosophy*: 379-399, p. 381.

> … by adding one or several more laws which contain the predicate and connect it with other terms available in our language. … In the case of the predicate 'soluble in water' we may perhaps add the law stating that two bodies of the same substance are either both soluble or both not soluble. This law would help in the instance of the match; it would, in accordance with common usage, lead to the result "the match is not soluble," because other pieces of wood are found to be insoluble on the basis of the first reduction sentence. (Carnap 1953, p. 56)

Fair enough. What if no piece of wood *whatsoever* had ever been put into water? Carnap admits that if some entity happens to consist of a substance such that no particular of that substance has ever been subjected to the test-condition for solubility, then we seem to be able to attribute neither the predicate nor its negation to it, which, again, is unacceptable. His remedy here is to claim that in such a case we need to state still further laws:

> These laws do not have the conventional character that definitions have; rather are they discovered empirically within the region of meaning which the predicate in question received by the laws stated before. But these laws are *extended by convention* into a region in which the predicate had no meaning previously; in other words, we *decided* to use the predicate in such a way that these laws which are tested and confirmed in cases in which the predicate has a meaning, remain valid in other cases. (Carnap 1953, pp. 56-57; my emphasis)

In other words: if it is the case that no piece of wood has ever been placed into water, then of course we cannot infer of this wooden match that it must be insoluble because other things made of wood are. We have, however, laws about other substances and their solubility in water, which have been tested by instances of dissolving in water, that may be projected beyond their range of confirming instances to also cover cases of submersion of wood. Carnap, unfortunately, gives little further detail on this process of extension, except by noting that it is "conventional." This is quite an elliptic, but important, claim—yet it is often ignored. For example, Mumford 1998, pp. 60-61, although reassuring the reader that he intends to discuss the empiricists' case in the most favourable light possible, ignores this passage and joins Pap 1963, p. 561, in falsely attributing to Carnap the rather unsatisfactory view that if a particular consists of a substance that has never been subjected to the test-condition for solubility, then we can call the particular neither soluble nor insoluble and the mere question whether it is, is itself meaningless. In fact, Carnap holds the more plausible, if undeveloped, view adumbrated in the quote above.

What sort of thought-processes, if any, could the relevant "convention" be based upon? The decision whether it is appropriate to extend applicability of a given law might, for instance, plausibly be based on an informed judgement concerning the *similarity* of the substances involved. If a given substance has, say, a molecular structure roughly isomorphic with that of another substance that has been observed to dissolve in water, we may feel justified in "extending" the relevant law in this case, and attribute solubility to the former as well. This would be an instance of inductive inference by analogy: first, we observe that

Some $a_1 \ldots a_n$ share the properties $P_1, \ldots P_n$

($a_1 \ldots a_n$ = objects or substances of some kind; P_n = the property of being soluble), where the fact that $a_1 \ldots a_n$ are soluble would be established by Carnap-style reduction sentences together with the plausible law that things of the same substance share at least some of their dispositional properties (namely those which are a direct consequence of being made of the substance in question). Then we also observe a given particular *b* of a different substance, which displays all but one of these properties,

b has $P_1 \ldots P_{n-1}$.

Through induction "by analogy" we finally infer that

b also has P_n.[63]

This is not the only way, however, in which we might ground the "convention" Carnap spoke of. Scientific disposition ascriptions to substances in remote space-time regions—e.g. to chemical elements shortly after the Big Bang, or under conditions to be found only at the center of certain stars, etc.—are likely to be based not on direct comparisons with existing elements and conditions. In such cases, the decision to extend a given law's applicability to a particular might be based on con-

[63] For proposals to extend Carnap's classic system of inductive logic to include induction by analogy, see Skyrms, B. (1993). "Analogy by Similarity in Hyper Carnapian Inductive Logic" *Philosophical Problems of the Internal and External Worlds*, Pittsburgh, University of Pittsburgh Press; Kuipers, T. A. F. (1988). "Inductive Analogy by Similarity and Proximity" *Analogical Reasoning*. D. H. Helman, Dordrecht, Kluwer: 299-313; Niiniluoto, I. (1988). "Analogy and Similarity in Scientific Reasoning" *Analogical Reasoning*. D. H. Helman, Dordrecht, Kluwer: 271-298; also Russell, S. (1988). "Analogy by Similarity" *Analogical Reasoning*. D. H. Helman, Dordrecht, Kluwer: 251-269.

sequences deduced from our background knowledge of other laws and principles. It is a fact of scientific life that the assumption that a given law applies outside its tested domain must be made under conditions of empirical unconfirmability and, sometimes, theoretical uncertainty. For instance, very little is currently known about the conditions in the lower regions of Jupiter's atmosphere and the behaviour of the prevailing gases under these conditions, because the temperature and pressure in this place of the universe far exceed anything reproducible in the laboratory. Jupiter's core represents a type of environment unfamiliar to us. Disposition ascriptions to particulars in this space-time region must be made on the basis of assumptions about how our current knowledge of the behaviour of gases under high pressure and temperature transfers to this sort of environment. We would construct a model of Jupiter's environment, i.e. a simplified representation of the physical system Jupiter and, inevitably, we do not know *a priori* that our model would include all parameters potentially relevant to the ascription. There is (so far) no way to go and find out directly. Insofar as model construction is influenced by such theoretical desiderata as simplicity, practical desiderata such as mathematical tractability, and meta-theoretical desiderata such as unification, Carnap's use of the term 'convention' in this context does not seem entirely unjustified. (For further discussion of this topic, see Chapter 3).

We need to consider a further case before we leave Carnap. What if the relevant test condition 'placed in H_2O' for solubility never obtained for *any* substance whatsoever, say because H_2O does not exist? *Only for this case* does Carnap's account indeed entail that 'soluble (in H_2O)' would be without meaning:

> If on the basis of either *logical rules* or *physical laws* it can be shown that all points belong to [a class of space-time points for which neither the positive nor the negative test-condition for the dispositional predicate Q_3 obtains], … then neither Q_3 nor '$\sim Q_3$' is determined for any point and hence the given set of reduction pairs does not even partly determine the meaning of Q_3 and therefore is not a suitable means of introducing this predicate. (Carnap 1953, p. 60; my emphasis).

This claim is rather more modest, and far less obviously absurd from the point of view of our realist intuitions in this domain, than the positions sometimes incorrectly attributed to Carnap. It makes little sense to ascribe, say, the disposition to a patient of being allergic to a certain chemical compound XYZ, if that compound is physically impossible and does not exist in this universe. All medical doctors to whom the present author has put the question agree with Carnap on this count: the ascription is neither true nor false but *meaningless*, insofar as, in their view, it does not make sense to either affirm or deny the truth of 'Patient N.N. is allergic to XYZ'.[64] The question itself is perceived as a typical philosopher's conundrum. Nevertheless, realists about dispositions find this consequence of Carnap's account deeply unsatisfying, as we shall see shortly.

One might wonder why Carnap, having admitted that it may be a law that 'bodies of the same substance' share the same dispositions, did not chose to simply stipulate possession of a certain substance—or, as we might say today, membership of a certain *natural kind*—as the defining criterion for the possession of certain dispositions, rather than the manifestation of certain experimental outcomes, or satisfaction of 'operational tests'. According to dispositional essentialism, some of the dispositions at least of a natural property are essential to that property, as seems to be the case for 'being positively charged' and the corresponding disposition to attract negatively charged objects (cf. Handfield 2001). Why not argue, in a similar vein, that it is *essential* to some substances (chemical elements) to be soluble in H_2O?[65] We may suppose that from the point of view of an empiricist, this would of course have meant putting the cart before the horse: the only way we can establish whether a given body is made of a given substance (or, for that matter, whether a given property is instantiated by an entity), is by conducting experiments and observing the results. The notion of 'substance' is to be admitted into our scientific language by way of a corresponding predicate that has been introduced through the very same method as dispositional predicates.

2.1.2 Goodman

Goodman 1983 adopts virtually the same criterion as Carnap for the correct attribution of dispositional predicates. He introduces a dichotomy of predicates, 'manifest' vs. 'non-manifest' ones, and suggests as a distinguishing mark of manifest predicates that they apply only to actual things and have classes of actual things as their extension (Goodman 1983, p. 41). According to this criterion, the predicate 'burns', for example, is a manifest one, because if it applies to an object at time *t*, it

[64] The sample size of this opinion poll is 3, hardly scientific, but nevertheless indicative.
[65] Thus, Bird, A. (2001). "Necessarily, Salt Dissolves in Water" *Analysis* **61**(4): 267-274 argues that the very fact of being a NaCl molecule *entails* that it is soluble in H_2O.

denotes an actual event occurring at t of which the object is an indispensible part at t. Goodman suggests we solve the vexing problem of un-manifesting dispositions in very much the same manner as Carnap: suppose the predicate 'M' applies to a disposition-specific type of manifestation-event—in Carnap's terminology, an experimental outcome of a test condition—, as for example 'burning' for flammability. Suppose further that burning happens to never occur. Then, by specifying a suitable causal connection in terms of causal laws between (the application conditions of) predicate 'M' and the (application conditions of) "auxiliary" manifest predicate 'A', such that when 'A' but not 'M' applies to an object, we may attribute the corresponding dispositional predicate D (Goodman 1983, Ibid.). Carnap's test-condition is, in Goodman's terms, the applicability of certain non-dispositional predicates.

The problem of unmanifesting dispositions, for Goodman, is essentially the problem of how to *project* the unproblematic, for manifest, predicate 'breaks', which applies to a subclass of those entities that drop, into a wider class to which applies the dispositional predicate 'fragile' (in other words, of how to establish a causal law that links fragility with braking events). Graphically:

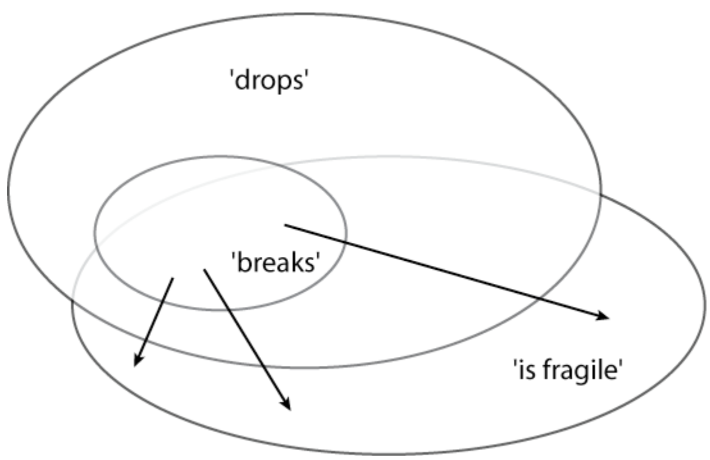

Fig. 6

Given its obvious proximity to Carnap's proposal, Goodman's approach is sometimes misunderstood as the attempt of a *reduction*—not of dispositional predicates to an accepted set of physical predicates used to characterize experimental outcomes, as in Carnap's case—but rather of dispositions to *possible events*. Many arguments fielded against Goodman are based upon this interpretation, and it must be conceded that this way of reading Goodman indeed comes to mind. For Goodman's usual examples for plausible manifest predicates are predicates applying solely to events ('burns,' 'brakes,' 'bends'), inviting the impression that disposition ascriptions according to

Goodman necessarily implicitly refer to events. D. H. Mellor, for example, claims that Goodman's answer to the problem of what to say about the fragility of undropping glasses is that undropping glasses must be assessed by their likeness to dropping ones, where 'likeness' consists in the fact that all fragile glasses share the applicability to them of a set of certain manifest predicates describing events (Mellor 1974, p. 111). This leads, according to Mellor, to the implausible consequence that whenever a disposition is not actually being displayed, some disposition-specific events are nevertheless "going on" (Mellor 1974, p. 112). Given that nothing whatsoever must be going on for an object to be the bearer of a disposition, Mellor concludes that Goodman's position is absurd.

Although it is correct to say that Goodman wants to solve the problem of umanifesting dispositions through an analysis of what links them to manifesting ones, Goodman 1983 does not in effect contain a commitment to an ontology of *events*. Goodman does not suggest, in particular, that for *a* to bear disposition D from t_1 to t_2, events of type *E*, of which *a* is in some sense an indispensable part, must occur during t_1–t_2. All Goodman says is that we can define 'flexible' in terms of manifest predicates if we find a manifest predicate that is related to 'bends under pressure' through causal principles or laws. Crucially, the manifest predicate does not necessarily need to be one applying to occurrences—it simply is, by definition, a non-dispositional predicate applying to 'actual things'. Goodman does adhere to the "empiricist" schema

$$D(x) \equiv \forall(x)\ (C(x) \rightarrow M(x)).$$

in the sense that for him, for x to bear disposition D, it must be a law, or at least non-accidentally true, that M(x) when C(x). But the question of x bearing disposition D is merely a question of a certain law being true of it, which implies nothing about any occurrence M(x) actually taking place.

Take electrical conductivity. We might be able to define '*a* is conductive' (where *a* = a piece of copper wire) if we found that some conjunction of predicates, perhaps

1. *a* is submerged in solution XY & X and Y separate
2. a is coiled around an iron bar & iron bar is surrounded by magnetic field[66]

---

[66] The two first predicate is from Quine, the second from Cartwright.

can be related to ‘*a* is conductive’ through law-like generalizations. The truth of $D(x) \equiv \forall(x)(Cx \rightarrow Mx)]$ does not entail that C(x) needs to obtain all the time, nor that it needs to be observable that it does—*idem* for M(x). It is therefore not correct to say that according to Goodman, something needs to be going on for a wire to be conductive, if this is to mean that something needs to be going on all the time for a wire to be conductive. Rather, Goodman says that the wire must be such that, by law, certain non-dispositional predicates are applicable to the wire under specified conditions. Obviously, this does not imply that the wire is not live while these conditions do not apply, for the wire could obviously still *satisfy* the law defining conductivity. In fact, the wire could satisfy the law even if these conditions never happen to apply. It is a law true of me at this moment that if I were to jump out of the window in front of me, I would accelerate downwards at roughly the speed of 9,8 m/s^2, but it is not true of me that I am currently jumping out of the window. Neither will I ever (I hope) jump out of a window, but this does not mean that I am not governed by the law of Free Fall. Similarly, the canonical test-condition for whether the wire is live is whether M(x) happens after C(x), where both C(x) and M(x) need to be described using manifest predicates, according to Goodman. But disposition-ascriptions do not, therefore, covertly refer to possible events, quantify over them, or rely on them in any other way. They do not even necessarily quantify over *actual* events: whether this is so depends on how one chooses to construe the ‘manifest’–‘dispositional’ distinction. Goodman stipulates, specifically, that manifest predicates are to be thought of as just those predicates, if any, which apply to *non*-dispositional things and adds that it is the dichotomy of the dispositional *vs*. the non-dispositional, rather than that of actual *vs.* non-actual *events*, which is essential to his proposal (Goodman 1983, p. 41n).[67] Disposition-ascriptions, on his account, *do* implicitly refer to whatever entities non-dispositional predicates can apply to (however we construe these), and to what is *nomically related* to whatever dispositional predicates can apply to.

[67] Here, Goodman does not at all endorse an ontology of *events*, nor, for that matter, does he claim that dispositional predicates denote *properties*. He merely talks nominalistically of individual ‘things’, leaving our options quite open concerning the nature of the primitive building-blocks of the universe, and the ultimate analysis of law-statements: “The predicate “burns” like the predicate “inflammable” is a word or label that applies to certain actual things and has the class of these things as its extension. Use of these predicates does not imply that they designate attributive entities; the predicates merely denote the things they apply to. A dispositional predicate, like a manifest predicate, is simply a term that applies to actual things; it need embrace no non-actuals in its extension.” (Goodman *Fact, Fiction, and Forecast*, pp. 41-42)

It is now obvious that Goodman's account suffers not from fallaciously trying to reduce dispositions to possible events, as frequently and erroneously stated, but rather from (a) staking too much on the availability of (strict) laws, and (b) threatening to make dispositions causally inefficacious. Take causal inefficaciousness first. Goodman says that a thing has certain dispositions and differs in that from others, because it is (actually) such that it satisfies a certain class of predicates, which are nomically related to disposition predicates. In other words, fragile glasses differ from non-fragile glasses because different laws concerning their non-dispositional properties are true of them. These laws correlate application conditions of non-dispositional predicates—without, to repeat, quantifying over *events*, just over whatever these predicates apply to, 'actual things'—, and so express actual differences between glasses. Whatever makes disposition-ascriptions true is hence whatever makes it true that certain predicates apply or fail to apply. When an object changes its dispositional properties, e.g. when it changes from being fragile to being non-fragile, different laws become true of it because certain non-dispositional predicates are no longer applicable to it, and others become applicable. This sort of picture has the drawback of inviting the conclusion that dispositions in themselves are causally inert, and that what really determines the causal powers of a disposition and explains an object's change in disposition is whatever determines the application-conditions of non-dispositional predicates, plausibly something itself non-dispositional. Of course, this sort of objection would not sway Goodman much, who, as a staunch empiricist and nominalist, would object to the whole notion of cause, or heaven forbid, causal power. I suppose that he would say that if his account makes no room for the causal powers of dispositions, the worse for causal powers.

The second problem might worry him more, however. It is simply that we do not know of any strict causal laws in connection with most dispositions. Even if there were a natural kind of all 'fragile' things, it would be very difficult to find a non-trivial description of it. What makes glasses fragile is very different from what makes bridges so, or stars, or marriages, etc. (Note, however, that the difficulty is the same for an account of dispositions in terms of the non-dispositional, "categorical", grounds of disposition-ascriptions, if the latter are meant to be micro-structural properties of substances. There does not seem to be anything in common to the micro-structure of all fragile things). The problem of the absence of laws is particularly serious in psychology: we do not currently know the causal laws, if any, which link predicates describing behaviour to disposition-words such as 'courageous,' 'thoughtful,' 'aggressive', etc. Goodman, clearly, has *strict* causal laws in mind. But if

Davidson 1970 and others are right, then there aren't any to be had in psychology, which would mean that we would have no definition of the meaning of 'courageous', etc., and no application conditions for it. Goodman, we should add, is aware of this. His *caveat* consists in pointing out that it may nevertheless be possible to define particular dispositional predicates through other means. "Abundant information" (what we would call today background knowledge) sometimes suggests a particular predicate 'P' that coincides in its application with the predicate 'flexible' , he says, in a way such that we may use 'P' as a *definiens* for 'flexible' even though we do not dispose of a theory formulated in terms of strict natural laws that links the two (Goodman 1983, p. 46). *Prima facie* at least, it would seem that this is exactly what we do with most psychological predicates, which, notwithstanding all our definitional problems, we use with considerable precision and sophistication.

It is worth mentioning at this point that notwithstanding the recent resurgence in realist approaches to dispositions and the concomitant demise of empiricism, some philosophers of science today still take it that dispositional predicates are defined by counterfactuals that link pairs of categorical predicates representing possible test conditions on the one hand, and corresponding outcomes (manifestations) on the other—and like Goodman they also assume that there is a close connection between laws and counterfactuals (cf. Liu 1999, p. 242n). The problem of strict laws vs. their non-strict counterparts, *ceteris paribus* laws, shall further exercise us in Chapter 3. It is, I believe, a central problem for any theory of dispositions, whether empiricist or realist.

2.2 Ontological Realism

Realist intuitions concerning the intrinsic independence of dispositions from any kind of manifestation, test condition, manifest property, or "predicate applicability" remain strong, notwithstanding Goodman's insights. In fact, the difficulty with disposition predicates whose test conditions never obtain seems unsatisfactorily resolved, by both Carnap and Goodman. For example, if the test conditions for a given disposition are physically impossible to realize (because the disposition is ascribed to an idealized entity in a theoretical context), then there will necessarily be no manifestation instances that could be described by manifest predicates, and no laws that could link these predicates to other predicates. To test the validity of such ascriptions we would need additional laws or postulates connecting "possible" manifest predicates—i.e. manifest predicates satisfied only by entities in possible worlds, a concept that would have been anathema to Goodman—with manifest predicates. We would, in other words, need an account of how to connect disposition ascriptions under idealized conditions to ascriptions under actual conditions (see Section 3.1). Finally, many philosophers want to resist not only the ontological reduction of dispositional properties to non-dispositional properties, events, or whatever-manifest-predicates-apply-to, but they also refuse to allow that we so much as *understand* the former in terms of the latter. To them, even though we frequently use conditional sentences to account for the meaning of disposition predicates, these are merely "inarticulate gestures" towards the way we really understand disposition ascriptions, namely as realist attributions of an actual, real, fully present property. The question is, of course, whether realist construals of dispositions in these terms can do better—a question that shall occupy us presently.

C.B. Martin and John Heil are amongst those who hold the view that a proper construal of dispositions shows that disposition ascriptions ought not to be reduced to, understood, or analysed in terms of conditional statements.[68] Dispositions are actual, real, properties and as such have nothing to do with potentialities and modalities (or, for that matter, predicate applicabilities subsumable under lawlike generalizations). Any traditional empiricist analysis of the form

$$D(x) \equiv \forall(x)(C(x) \rightarrow M(x))$$

must be wrong, in their eyes, for whether any x has disposition D is not dependent in any way upon manifestation events M, or stimulus conditions C. Philosophical accounts of dispositions have, for too long, laboured under a "common misunderstanding".

[68] Martin has argued this in many places. Here, we shall be most concerned with Martin, C. B. and J. Heil (1997). "Rules and Powers" *Language, Mind, and Ontology*, Cambridge, Blackwell, **12:** 283-311.

2.2.1 What Do Statements About Dispositions Mean?

The misunderstanding, as Martin and Heil see it, originates precisely in what we identified above as one of the pre-theoretical intuitions about dispositions, namely the widespread view that statements ascribing a disposition to a given object 'mean the same as' or entail certain conditional statements. But Martin and Heil point out that although the statement

a is fragile

is generally thought to mean the same as

If *a* were stressed, *a* would break,

'*a* is fragile' does not actually *entail* 'If *a* were stressed, it would break', and really it cannot be construed as meaning the same either. We should ask ourselves why we should have thought this in the first place: everyone knows that an object does not break every time when stressed! Moreover, we also know that even if a particular stressing event were on its own sufficient to break *a*, external circumstances and conditions might nevertheless be such that the event is not followed by *a*'s breaking. For example, a stressing of a particular glass normally sufficient to break the glass might take place under very high temperature. In that case, the heat would make the glass flexible enough to enable to absorb a large part of the shock energy through deformation. Object *a* deforms, but it does not break. Thus, D. H. Mellor reminds us that 'If *a* were stressed, it would break' can only be true for specific types of stressing event and specific initial conditions, as well as for specific types of object (Mellor 2000, p. 759). Indeed, similar qualifications are needed for the conditional analysis of *any* disposition ascription, because the way in which a given disposition actually displays (and whether it does manifest at all) is always dependent upon contingent factors.

Some authors, though not inclined to accept the empiricist theory of dispositions, are determined to maintain some sort of conditional analysis, and hold that given sufficient sophistication, a theory employing subjunctive conditionals will do the job. According to Mellor, for example, the rough outlines of an account of the meaning of '*a* is fragile' could be:

> x is fragile iff [if x is of kind K and is relatively suddenly and lightly stressed in way W_K, x would break] (Mellor 2000, p. 759)[69]

The phrases 'relatively suddenly and lightly' and 'way W_K' attempt to encode the fact that dispositions may be more or less 'generic,' as Mellor puts it.[70] 'Fragile' is very generic: as already pointed out above, to be fragile for a bridge is an entirely different thing than for a glass or a cease-fire, and whatever it is that makes a bridge fragile (its 'physical basis', if any) is likely to be different from whatever it is that makes a cease-fire fragile (its physical basis, if any). Accordingly, different kinds of stress are necessary to break one and the other, notes Mellor. Additional complications are to be expected with the notion of 'kind' and the vague terms 'relatively' and 'lightly', and Mellor acknowledges that the sort of analysis he advocates requires further fine-tuning.

But Martin and Heil harbour fundamental objections against the whole conditional approach, objections which would stand even if the minutiae of such an account were to be satisfactorily worked out one day. They believe that no matter how many refinements we introduce, the suggested *analytic link* between a given disposition ascription and its correlate counterfactual conditional(s) can always be shown to brake down under specific conditions. These are conditions under which the very factor that normally makes a disposition's manifestation possible (i.e. condition *C*) actually impedes it. Martin's well-known example involves a hypothetical device—or, rather, a class of functionally equivalent devices—such that whatever object *a* with dispositional property *D* the device is attached to, it has the following effect: *a* loses *D* when and only when *C* obtains (Martin and Heil 1997, p. 3). For example, an 'electro-fink' is a machine which

> ...can provide itself with reliable information as to exactly when a wire connected to it is touched by a conductor. When such contact occurs the electro-fink reacts (instantaneously, we are supposing) by making the wire live for the duration of the contact. In the absence of contact the wire is dead. ... In

---

[69] For another recent defence of the conditional analysis of dispositions, see Malzkorn, W. (2000). "Realism, Functionalism and the Conditional Analysis of Dispositions" *Philosophical Quarterly* **50**: 452-69.

[70] Mellor reports the distinction between 'generic' dispositions, i.e. dispositions that manifest themselves in many different objects under many different manifestation stimuli, and 'specific' dispositions that manifest only in one type of object under one type of stimulus, as due to Ryle *The Concept of Mind*, the grandfather (within contemporary Anglo-American philosophy at least) of explicitly dispositional analyses of the mind.

> sum, the electro-fink ensures that the wire is live when and only when a conductor touches it. (Martin and Heil 1997, p. 3).

It could be true of a copper wire connected to an electro-fink that 'if the wire were touched by a conductor, electrical current would flow from the wire to the conductor,' but the truth of that counterfactual would not suffice to establish the truth of the disposition ascription 'the wire is live.' The *physical* possibility of electro-finks and of similar gadgetry is not at issue here. An omnipotent agent could fulfil the same role, allows Martin, and omnipotent agents seem even less physically possible than electro-finks. It is mere logical possibility, which is meant to suffice to establish the desired conclusion that disposition ascriptions cannot mean the same as counterfactual conditionals, no matter how much the latter are propped up.

The concept of a 'finkish disposition'—a disposition which, although it persists in its bearer during the period when its manifestation is not called for, cannot manifest due to the external circumstances of its bearer—may be put to work in the context of Kripke's scepticism. Suppose, write Martin and Heil, agents Don and Van each acquire a different rule for the use of the symbol '+.' (Martin and Heil 1997, p. 285). Don learned the *plus* rule, whereas Van acquired *quus*. Suppose further that Van inhabits a world where a Cartesian-style *malin génie* ensures that on every occasion on which Van intends to compute *quus*, he errs and produces results that correspond to the prescriptions of the *plus* function. In such a case, ... it could be true of Van that were he to consider any two integers, he would 'sum' them in a way that perfectly conforms to the plus table (Martin and Heil 1997, p. 285). Through the intervention of the evil spirit, Van effectively acts 'in accord' with the *plus* rule, as Martin and Heil say, but he does not act 'on' the rule. In other words, notwithstanding the truth of the above counterfactual, it would not be true of Van that he computes the *plus* function. (Not, at any rate, on a description which takes into account his intentional mental states). Clearly, we must say that Van, although interfered with by the demon in such a way that his behaviour is identical to Don's, still computes and intends to compute *quus*, and continues to bear the disposition to do so. His disposition is simply *finkish* in the sense that, although he does bear the disposition, whenever circumstances are such that its manifestation is called for, the latter is somehow rendered impossible by the demon.

According to Martin and Heil, recognition of the possibility of a disposition's finkishness not only represents a solution of the vexed problem of how to explain error (in other words, the normativity problem—a claim we shall not examine here), but it also points the way towards a realist reply to the charge that dispositions cannot account for the infinite application of a rule. For, once we use a less "crude" technique for identifying a given agents' commitment to a particular rule with his dispositional make-up, we will realise, the authors hold, that even though Don's acquisition of the *plus* rule amounts to his acquisition of a particular disposition, this does not mean that

> … Don will manifest it on every occasion in which its manifestation is called for. Don's overall dispositional state could be such that, on particular occasions, the manifestation of the disposition constituting his possession of the plus rule is blocked or inhibited. Don and Van do differ, then. The dispositional state constituting Don's grasp of the plus rule is intact, although, at *t*, its manifestation is thwarted. Van, in contrast, possesses no such dispositional characteristic. (Martin and Heil 1997, p. 291)

Consideration of the case of finkish dispositions forestalls the 'elementary confusion to think of unmanifesting dispositions as unactualized *possibilia*.' (Martin 1994, p. 1).

Bird 1998 invokes a similar conundrum, "anti-dotes": if circumstances were such that the very dropping of a glass onto a floor softens the (normally hard) floor so that upon impact the glass does not break, then a glass could be fragile without the corresponding counterfactual being true. The floor's softness would constitute an *antidote* to the glass's fragility. Indeed, real-life antidotes do just this, they neutralise certain specific dispositions of poisonous substances. The point made by Bird is essentially the same: dispositions are to be thought of as actual—indeed as actual as anything can ever be—even when circumstances such as the presence of various sorts of defeaters such as electro-finks, evil spirits, and anti-dotes, impede their manifestation. Defeaters need not be outlandish sorts of things, comparatively more pedestrian examples are at hand. Take a possible world containing salt but no water: (Martin and Heil 1997, p. 295). NaCl has the dispositional characteristic of being soluble in water, yet in that world any potential manifestation of that characteristic is inhibited by a feature of the circumstances of each NaCl molecule, namely the total absence of H_2O. Plausibly, this circumstance does not imply, *eo ipso*, that salt would not be soluble in such a world, and that inhabitants of that world (or we, for that matter) would not be justified in making the disposition ascription. Martin and Heil therefore disagree on this one with Carnap and the members of the medical community we have polled (see Sec. 2.1.1).

Generally, to be a realist about dispositions is to believe that a disposition ascription tells us how the subject of ascription *is* right now in actual circumstances *X,* rather than saying what it would do in possible circumstances *Y*. What it would actually *do* in circumstances Y is a matter of further facts obtaining or failing to obtain. Disposition ascriptions, on a realist account, do not refer to possible manifestations, nor to any sort of event or potentiality at all, but ascribe actual properties. Consequently, whether these manifestations are physically possible or not must be immaterial to the ascription.

2.2.2 The Metaphysics of Dispositions

Martin and Heil thus advertise a notion of dispositions that is meant to be both more 'fine-grained' and 'robust' than the empiricist one: it is possible for an object to bear a disposition the *manifestation* of which has been permanently "blocked" by some factors impinging on the object, yet whose *existence* is unaffected by this. The blocked disposition remains deep-seated in, or possessed by, the object, and if whatever factor inhibiting the disposition's manifestation is removed, the disposition returns to its previous uninhibited state. It would be false to say that the disposition is, at that moment, *re*-acquired by its bearer. For possession of the blocked form of a disposition *D* is sharply to be distinguished from the loss or lack of *D* (Martin and Heil 1997, p. 291). It is important to note that although their inaugural example is, even by philosophy's standards, quite contrived, the realist claims that Martin and Heil proceed to make attempt to translate a fairly common realist intuition about dispositions. Take, for example, medical conditions such as transient aphasia. Patients with localized brain injury due to strokes, accidents, infections, etc., are sometimes observed to partially loose their capacity of production and/or comprehension of speech and the ability to read or write (*aphasia*). However, a large proportion of those who initially show symptoms of aphasia recover completely within the first few days. There is a strong case for claiming that transient aphasia does not cause patients to loose the relevant dispositions, which are then miraculously reacquired, but rather that it *inhibits* the dispositions' manifestation for a certain time.[71] .

Conceptualising dispositions in this way requires acknowledgement that dispositions typically occur in complex clusters rather than in isolation, clusters which

---

[71] Chomsky *Knowledge of Language: Its Nature, Origin, and Use*, pp. 9-10, makes just such a claim. Note, however, that this example does not involve the trickier case of a disposition whose manifestation is physically impossible

allow for various kinds of interaction between dispositions that influence the degree to which any individual disposition can manifest itself. Thus, the human brain is, according to Martin and Heil '... a complex object, one the dispositional structure of which is staggeringly intricate. If we consider the brain as a dispositional array, we can see it as possessing sharply focused dispositional structures, organized collections of dispositions that manifest themselves in inhibiting or enhancing other dispositions, and dispositions for the acquisition of further dispositions.' (Martin and Heil 1997, p. 294). The idea of the brain as a dispositional array with dispositions embedded in each other and collectively determining the overall behaviour of the system, is reminiscent of what Cartwright 1997 calls a *nomological machine*.[72] This is a permanent arrangement of capacities, and it is *nomological* because its repeated 'operation' (i.e. the capacities' repeated manifestation) in a stable environment gives rise, according to Cartwright, to the sort of regularities we describe in our scientific laws. Cartwright is not the only contemporary philosopher of science who thinks that it is time to finally put our positivist scruples behind us and to endow capacities, or dispositions, with a scientific status (see , …). Disagreement persists on whether this ought to initiate a fundamental rethink of the status of laws of nature. Cartwright 1989 argues that laws of nature are but the by-product of the operation of capacities, they are the generalizations we are sometimes lucky enough to stumble upon if nature in a given domain or spatio-temporal region happens to be kind, and regular, enough (cf. also Mumford 1998). Others are more conservative and would like limit the revolution to admitting dispositions as the relata of scientific laws (Liu 2001b). Be that as it may, all of these philosophers conceptualize dispositions much like Martin and Heil, als self-contained entities that combine in various ways, and that can subsist even when circumstances are not favourable for their manifestation, i.e. even in a permanently inhibited state.[73]

[72] Cf. also Cartwright, N. (1989). *Nature's Capacities and their Measurement*, New York, Clarendon Oxford Press. The reminiscence is far from coincidental, of course, as all concerned authors are realists about dispositions/capacities, and consider them as ontologically basic entities with an irreducibly modal character. The *caveat* "in a stable environment", or what I take to be its equivalent, "everything else being equal", will be at the centre of our attention in Section 3.1.

[73] Liu, C. (2001b). "Laws and Models in a Theory of Idealization", http://philsci-archive.pitt.edu/documents/disk0/00/00/03/63/index.html. (accessed 29/01/2002), for example, holds that many natural laws relate dispositional properties, and that we should hence be prepared to be realists about them. However, it is also true that dispositions often combine and superimpose in such a way that conditions are not always ideal for their "display", or manifestation. Even here, he explains, we should assume that the dispositional properties which '… the laws relate are instantiated *fully* in those non-ideal circumstances and thus the laws obtain. What is not always true for such laws when they are co-instantiated with other such laws is that they can show themselves in their full categorical display.

This means that any given disposition is to be considered a potential *disposition partner*, among infinitely many alternatives, for a possible mutual manifestation with other dispositions. For example,

> The very same dispositional state of water can have dramatically different mutual manifestations, depending on its reciprocal partners: consider water sprayed on flaming oil and water sprayed on flaming timbers. ... this, perhaps infinite, range that any disposition has for different kinds of alternative reciprocal disposition partner, actual or nonactual, for different kinds of mutual manifestation, is wholly present in any time-slice of a disposition (Martin and Heil 1997, p. 295)

The size of the infinite set of possible manifestations that any given disposition is a disposition *for* (is selective for) is limited by the fact that a dispositional state cannot be for just *any* sort of manifestation. As Martin puts it elsewhere, a square peg does not have the disposition to fit into a round hole in precisely the same way as a round peg does. Dispositions are irreducibly modal, they ‘may be seen as programmes for infinities of manifestation with limited scope. … The *limits* of directedness are *set* by the still greater infinity of manifestations the disposition is *not for* as *well* as the manifestations it *prohibits*’ (Martin and Heil 1997, p. 517). A given dispositional state *D* thus determines its set of alternative partners for manifestation in virtue of excluding the still larger set of manifestations it is *not* for, or “prohibitive against.”

The modal nature of a disposition, its *projectivity* or *directedness*—which obtains of course without the benefit of a causal link between a given dispositional state and its set of possible manifestations—is precisely what Martin and Heil think endows a dispositional state with content. As Martin explains, ‘There is a parallel between the *content* of a mental state and the ‘*what for*’ of a non-mental disposition. … Just as a belief needs *content*, namely, what would follow the ‘that’ in ‘belief that…’ so a disposition needs a *what-for*, namely what would follow the ‘for’ in ‘disposition for…’, e.g. ‘disposition *for* dissolving in H_2O’ (and *not aqua regia* or glass)’ (Martin 1993, pp. 516-17). Just as a mental state may be directed towards unicorns, a disposition can be directed towards what is non-existent. A particular substance may, for example, have the disposition to dissolve in a particular solvent, which does not exist in nature, but which might be manufactured if necessary. Martin and Heil hence share the view of U.T. Place, who has argued that the independence of the truth of a dispo-

Therefore, *idealized models (or ideal conditions) are necessary for the laws to show their full categorical glory*.’

sitional statement from actual manifestations or actual states of affairs shows that a disposition is something that "points" beyond itself to what does not exist, thereby displaying a kind of intentionality.[74]

Reference to a disposition's built-in projectivity or intentionality, then, is the way to go when dealing with the case of meaning *plus* by '+'. According to Martin and Heil it is vital, however, to distinguish between the disposition itself, which constitutes mastery of the rule, and what they call its accompanying 'reciprocal capacities.' A reciprocal capacity of the disposition to use '+' as denoting *plus* might be, they say, the capacity to entertain thoughts involving applications of the *plus* rule, or the capacity to write down the results of applying the rule (Martin and Heil 1997, p. 301). This distinction allows the dispositional theorist to acknowledge that the counterfactual

If given two integers i_1 and i_2, Don could work out their sum on paper

will never be true of Don for *any* i_1 and i_2. He does not take this fact as implying that Don has not mastered the addition rule. Rather, Don fails because his relevant reciprocal capacities are not up to the task (Martin and Heil 1997, p. 301). (This idea, incidentally, would be a bone of contention with Wittgenstein. To the latter, what is viewed as 'reciprocal' and hence inessential to mastery of a rule here, is our sole "criterion" for mastery: no correct ascription of mastery of a rule without some sort of ability of actually applying it that counts, within the relevant community of rule-followers, as an ability to apply that rule). Martin and Heil believe that Don's limitation with respect to addition results from limits inherent in capacities concomitant with his disposition to add, not in the latter disposition itself. Don can therefore truthfully be said to have acquired the *plus* rule, and hence a disposition to add, without necessarily being able to manifest that disposition on every occasion. (In actual fact, for the *vast majority* of 'stimulus' or 'test' conditions *C* under which we would expect the disposition's manifestation, e.g. queries of the form 'x + y = ?', Don will be unable to manifest it.) This "simple point" is obscured, according to Martin and Heil, by the widespread failure on the part of philosophers to make the distinction between mastery of a rule on the one hand, reciprocal dispositions that determine an agent's capacity to display behaviour that counts as a manifestation of his mastery of the rule

[74] For the latest exposition of that view, see e.g. Place's contribution to Armstrong, D. M., C. B. Martin, et al. (1996). *Dispositions: A Debate*, New York, Routledge.

on the other, as well as the finiteness of the latter, and his circumstances (Martin and Heil 1997, p. 302).[75]

This distinction echoes another that is frequently invoked in connection with Noam Chomsky's famous performance-competence dichotomy.[76] Chomsky distinguishes a speaker's internally represented grammar, or his knowledge of language (= his competence), from other mental subsystems, such as the brain's 'parser,' that interact with it when the internally represented grammar is put to use in understanding and producing sentences of the relevant language (= his performance). According to Chomsky, the speaker's limitations as a finite agent are due to limitations in these latter subsystems, not to limitations in his competence (generally, factors believed to affect performance include attention, stamina, memory, and even my beliefs about my interlocutors). Hence the not uncontroversial claim that an English speaker *qua* English speaker has the ability to parse and assess the grammaticality of any English sentence of *any* length, but for insufficient resources available to the speaker's parser, e.g. insufficient short-term memory. The idea at work in the linguistic case is the same as in our present context, namely that 'mastery' of a rule, or of a set of rules governing a certain practice, is instantiated in an agent entirely independently of the existence and performance of what Chomsky calls sub-systems, and Martin and Heil call 'reciprocal capacities.' Martin and Heil put it as follows:

> We are supposing that an agent's possession of a rule might be constituted by his dispositional condition. Imagine, now, that A's possessing the plus rule is constituted by some component, P, of that condition. Similarly, B's possession of the quus rule is constituted by component Q of B's dispositional makeup. It could easily be the case that P and Q situated in less limited environments, would, under appropriate triggering conditions, manifest themselves in a way that the divergence ... would show itself. We could imagine transplanting' P and Q, or simply supplementing A and B in appropriate ways. (Compare adding memory to a computing machine to enable it to compute larger or more complex numbers). (Martin and Heil

---

[75] Unfortunately, Martin and Heil do not provide us with a clear distinction between dispositions and reciprocal capacities. It seems that reciprocal capacities are just *other* dispositions we also happen to have. Martin and Heil here echo Simon Blackburn, who claims that the sceptic's main oversight is that a correct dispositional theory of meaning *plus* by '+' would take into account further dispositions *surrounding* our disposition to give answers (Blackburn "The Individual Strikes Back", p. 290). Even though the latter is indubitably finite, these other dispositions make the judgement that we follow the plus-rule the only possible judgement. Blackburn concedes, however, that this does not exclude inductive scepticism about the concept of a disposition (Ibid.).

[76] See Fodor 1985, p. 154; Stich 1990, p. 185; Patterson, S. (1998). "Competence and the Classical Cascade: A Reply to Franks" *British Journal for the Philosophy of Science* **49**(4): 625-636, at pp. 631-634. For a discussion of dispositions *ceteris paribus* in terms of the notion of competence, see *infra*.

> to compute larger or more complex numbers). (Martin and Heil 1997, p. 300)

It is of course rather plausible that there is nothing qualitatively different about, say, the mathematical knowledge and cognitive capacities necessary to add 2 and 2, and the knowledge and capacities necessary to add two 1000-digit numbers, but for a series of purely quantitative issues concerning memory, concentration, the most efficient algorithm to use, etc. It is also tempting to conclude that whatever it is about me that makes it the case that I have learned how to do *addition* and not *quaddition* somehow projects beyond my finite capacities:

> A dispositional component of a finite system can 'project' beyond the capacities of the system to enact or manifest what is projected. … If a disposition's projective 'direction' is intrinsic to it, it is present even if it is never manifested, or, owing to limitations of the system to which it belongs, its manifestation is not physically possible.

It seems indeed that once we have learned how to do the calculation for small numbers we are, at least in principle, ready to go to do it for all numbers—all depends on what the phrase 'in principle' is made out to mean. I submit that in this context, it means roughly what the phrase "*ceteris paribus*" means. For instance, '*ceteris paribus*, if demand rises and offer remains constant, the price will rise', is usually construed as saying that the sort of functional relation, which according to the law obtains between the quantities mentioned, obtains under the condition that no factors other than those mentioned in the law itself interfere. In other words, demand, offer, and price are treated as an *isolated system* for the purposes of establishing the lawlike regularity. This does not mean that they need to be in actual fact—in fact they never are in reality. The *ceteris paribus* clause thus serves to save the law from instant falsification by a world where price, demand, and offer never constitute an isolated system. Similarly, the claim that possession of the plus-disposition equips its bearer to add huge numbers must be hedged by something equivalent to a *ceteris paribus* clause. If no additional factors entered the picture that inhibit me from manifesting the disposition, then I could add these numbers; but I cannot, which shows that additional factors do enter the fray, and that "everything is not equal". In Martin and Heil's terminology this is the claim that the relevant disposition, although present and "ready to go" in the bearer, is never in fact isolated, but always part of a dispositional structure that consists of a multitude of other (finite) dispositions and reciprocal capacities, which for a large part inhibit manifestation of the plus-disposition.

2.2.3 Pregnant Spinsters and Unwanted Children (Epistemological Worries I)

Commenting on the history of the notion of 'disposition' in 20th-century philosophy, D. H. Mellor writes: 'Dispositions are as shameful in many eyes as pregnant spinsters used to be—ideally to be explained away, or entitled by a shotgun wedding to take the name of some decently real categorical property. It is time to remove this lingering Victorian prejudice. Dispositions, like unmarried mothers, can manage on their own. They have been traduced, and my object here is to restore their good name.' (Mellor 1974, p. 157), As shameful as pregnant spinsters dispositions in philosophy undoubtedly have been, and discriminated against by self-righteous philosophers—from the mechanists on all the way to modern empiricists—they admittedly were. Yet, the brave new world of respectability ushered in by contemporary realism about dispositions does not entirely liberate them from their oppressive past. In fact, it comes back to haunt them in the form of an unwanted child, by the name of 'Epistemology.' Or so I shall argue in this Section.

The sceptic's objections to Martin and Heil's proposed solution are likely to be of the number of at least two. The first is "a natural one for the sceptic to raise," as Martin and Heil are quick to acknowledge, namely that they have not answered the question 'How do you *know* that you are following the rule for *plus* rather than the rule for *quus*?' (Martin and Heil 1997, p. 307). The worry is that the picture offered fails to illuminate the process, if any, through which we find out about the precise nature of a given disposition's 'projectivity,' and therefore about *which* specific dispositions a thing bears. It does not help that the projectivity of a disposition is explicitly ruled out from being grounded in any sort of *causal* relation. Studying, say, the *causal role* of an agent's putative dispositional states therefore looks rather hopeless if what we are after is criteria of individuation (especially, and particularly obviously so, in arithmetical cases). Once it is acknowledged that a dispositional state projects towards its possible manifestations in not quite the same manner in which a cause is related to its possible effects, the sceptic will legitimately question whether anything remains in the real world which *makes it the case* that a disposition D is connected to one set of possible manifestations with appropriate ‘mutual manifestation partners,’ rather than to another.

The second worry is that even if we grant such a selective and thereby content-determining connection between a given disposition and its unique set of reciprocal partners (what Martin and Heil call a disposition’s ‘disposition line’), the theory does not really *explain* it—in the sense in which we explain an event or a fact by, say, subsuming it under a law, or by indicating the underlying mechanism that brought it about. On the face of it, projection is in need of explanation, however. A dispositional state, in the mind of Martin and Heil, is something real and actual; the disposition line to which a disposition projects, on the other hand, is an abstract object, an infinite set containing indefinitely many non-existing objects or events (manifestations). This means that if a disposition uniquely determines its disposition line, as Martin and Heil claim it does, then a disposition is an infinitary object in Boghossian’s sense. One should think that for someone in the business of arguing that we should accept dispositions as real objects, as things in the world, explaining how disposition lines, i.e. manifestations, are related to dispositions would be of the utmost theoretical importance. It is certainly of the utmost importance for countering the Sceptic—after all, Martin and Heil explicitly argue that it is *in virtue of* their being correlated to different disposition lines that the dispositional states of agents Don and Van differ. Yet they are silent on the process through which a unique disposition line is correlated with a unique dispositional state. Are dispositions embodied *functions*? How could they fulfil the role of a function? Again, we have come upon the quite mysterious nature of anything purported to be infinitary.

Concerning the first difficulty, Martin and Heil seem to have a quick way out. The sceptic, they note, grounds his challenge on the question of evidence, namely the evidence one could have that one is following a particular rule. The challenge gets off the ground precisely in virtue of the fact that it is not obvious what sort of evidence I might have for my possessing the *plus* rule rather than the *quus* rule. However, this whole approach to the matter is ‘wrong-headed,’ they say, because ‘The skeptic assumes without argument that, if you know that you are following the plus rule, you do so straightforwardly on the basis of evidence. But why should we agree? You know that you are now awake, but not on the basis of evidence about your physiological makeup.’ (Martin and Heil 1997, p. 308). First-person knowledge and the corresponding authority concerning claims about inner states are indeed not generally assumed to be based on evidence in the classical sense. Martin and Heil believe that it is based on something rather different, namely what they call ‘cue manifestations:’

anyone who has mastered a rule *R* bears, according to them, both a disposition *D* constituting mastery of that rule, as well as a collateral disposition *D´* to 'appreciate the rule *R* and its lack of limit as the rule he is following.' (Martin and Heil 1997, p. 302). This collateral disposition in turn gives rise to cue manifestations of *D,* which are claimed to be at the origin of the agent's sense of being ready to "go on" in accordance with *R*. A cue manifestation is what allows us to assess our capacity for a task without having to complete that task, or even to begin it (Martin and Heil 1997, p. 297; and Martin's contribution to Armstrong, Martin *et al.* 1996). A cue manifestation, in other words, is the sort of inner experience that tells me that I can jump across this pool of water without getting wet. Crucially, this experience is only '... a reliable, though not infallible, indication of the presence of a particular dispositional condition. You might know, perhaps, that you are in a particular dispositional state on the basis of some cue manifestation. The latter need not function as evidence for you, however, it need only connect your belief about your current dispositional state to that state' (Martin and Heil 1997, p. 308).

Thus, the sceptic's demand for a fact that not merely shows that Don means *plus* by '+,' but also justifies Don's response as the right one, is too strong. Don may believe that he means *plus* on the basis of "evidence" that is not intersubjectively demonstrable, for cue manifestations are inherently private. Nonetheless he may be *justified* in doing so. Generally, it is just too strong, say Martin and Heil, to require that a justified believer 'also believe justifiably that the justifying conditions obtain, and perhaps believe as well that these conditions are in fact justificatory' (Martin and Heil 1997, p. 292). Justification does not, on this sort of view, require the agent to be able to come up with a justifying reason. Certain dispositional states of Don could be thought of as justifying an action they cause without their justificatory character depending on Don's explicitly appealing to, or even recognising, them as justificatory. For, 'An agent's actions or beliefs might be regarded as justified in light of some rule, R, if we understand the agent's acceptance of or commitment to R as constituted by a suitably deep-seated dispositional characteristic of the agent (together with an absence of defeaters), and his actions or thoughts as appropriate manifestations of that dispositional characteristic' (Martin and Heil 1997, p. 293).

Martin and Heil's account of what it means to be justified essentially follows 'inferential externalism' in epistemology. Inferential externalists are diametrically opposed to their internalist counterparts, who assert that for my belief in proposition P_2 to be my epistemic reason to believe in proposition P_1, a "connecting" belief is necessary in a third proposition P_3, that P_2 'speaks in favour' of the truth of P_1. The internalists say that without the connecting belief, P_2 could not *count* as my epistemic reason for believing P_1. In other words, one cannot be justified in believing that whatever makes P_1 true obtains *because* whatever makes P_2 true obtains, without further believing that another fact/state of affairs obtains to the effect that P_2 is a good reason to believe that P_1. From the internalists' point of view, this picture best accounts for our intuitive understanding of what it is to be justified in believing 'P_1 because P_2'—after all, what *else* could it mean to say that one's belief in P_2 is an 'epistemic reason' to believe in P_1?

Externalists object that we are lead into a vicious infinite regress: it is plausible to require that the connecting belief itself be justified by a further belief, for how could P_3 constitute an epistemic reason for believing that P_2 supports P_1 if there were no proposition P_4, the belief in which would constitute our epistemic reason for believing in P_3, and *ditto* for P_4? Their solution is to do away with the requirement of a connecting belief, and to replace it with an *inferential disposition*. Thus, to correctly infer a belief that P_1 from a belief that P_2, it suffices that P_1 and P_2 instantiate a (valid) inference pattern, and that my inferring from P_1 to P_2 is the effect of my bearing a disposition to make transitions which conform to that pattern. The threat of infinite regress is avoided by stipulating that one can be justified in inferring P_1 in virtue of the mere existence of an appropriate relation between P_1 and P_2, the existence of that relation allowing justification to be transmitted from the latter to the former. Similarly, Martin and Heil's stance is presumably intended to forestall any regressive move on the part of the Sceptic, who, after being presented with a putative fact justifying the relevant disposition ascription, might proceed to ask what justifies us in believing that *that* fact is indeed justifying, and demand a higher-order fact proving that it is justifying, etc. The position is therefore that Don's commitment to *plus* rather than *quus* is justified simply in virtue of his bearing a deep-seated disposition to mean *plus* by '+'. End of story.

Whatever the merits of externalism in solving the problems that beset internalism about justification,[77] it patently lacks the resources to solve our problem. Mar-

[77] The objection to externalism most relevant to our concerns is probably that if it is something *external* about my opinions that determines whether they are justified (e.g. the fact that they issue from a set of inferential dispositions with the right sort of causal history), then I am not, with respect to the majority of my opinions, particularly well placed to tell whether I should have them. After all, I do not know exactly what sort of processes have caused most of my opinions, and whether these processes have the properties necessary and sufficient to justify the opinions they cause. Although the justification of my beliefs is clearly not something over which I have first-person authority, any theory seems excessive that suggests that there is not at least a substantive sub-set of my beliefs for the justification of which I *need not* defer to experts. This is related to our intuition that for at least a substantive sub-

tin and Heil cannot simply argue that Don's response '125' to the query '67+56= ?' is justified *because* Don bears the deep disposition to make judgements of a sort that conform to the pattern prescribed by *plus*. This obviously misses the point. For the Sceptic may agree that there is indeed an 'appropriate relation' between the query '67+56= ?' and the response '125,' and further follow the externalist in allowing the conditional claim that *if* Don bore the disposition to add, he would *eo ipso* be justified in responding as he does. The sceptic simply holds that Don is not so justified precisely because he does not bear the relevant disposition. Rather, Don (like the rest of us) bears a member of $\{\Phi\}$, where '$\{\Phi\}$' is the infinite set of distinct dispositions the manifestations of which extensionally coincide with Don's (our) previous applications of '+' (cf. the problem of redescription). Both the relevant *plus*- and *quus*-dispositions are members of that set, and we are entitled to conclude, according to the sceptic, either that Don bears no determinate disposition at all—a conclusion Kripke seems to advocate when he comments that the paradox endangers our notion of meaning *tout court*—or to hold that even though we know that Don bears a member of $\{\Phi\}$, there is no fact of the matter *which one* he bears. In other words, there is simply no fact of the matter as to whether Don is justified or not, even on the externalist story.

The essential sceptical doubt begins at an earlier point than Martin and Heil surmise. For the sceptic denies Martin and Heil's initial premise that there is such a thing as a dispositional state that is selective for, or projects to, *one* particular mathematical function rather than another. In other words, the sceptic denies that dispositional states map one-to-one onto sets of possible manifestations (disposition lines), and thus denies that there can be dispositional states which bear a determinate content in the manner described. The Sceptic's reasoning is based on the premise that all my dispositional states are essentially finite, and denies that they could uniquely determine something infinite (a set of manifestations), i.e. that they could be infinitary. In Kripke's words

> The dispositional theory … assumes that which function I meant is determined by my disposition to compute its values in particular cases. In fact, this is not so. *Since dispositions cover only a finite segment of the total function and since they may deviate from its true values*, two individuals may

set of my mental states, I require no evidence to know whether I have them, and what they are. However, the Sceptic's challenge being what it is, it seems that no matter whether Martin and Heil choose epistemological externalism or internalism, their choice would appear to have no impact on it (see *infra*). I am indebted to Mark Textor for discussions on this subject.

> agree on their computations in particular cases even though they are actually computing different functions. Hence the dispositional view is not correct. (Kripke 1982, p. 32; emphasis mine)

The essential disagreement hence concerns the question whether a finite agent's dispositions really could be more than finite and, although they but cover a segment, "cover" in a different sense the total function concerned—namely by projecting towards it. Martin and Heil in effect *stipulate* that the very nature of dispositions is such that they can (see our citation *infra*), and take it as their axiom that dispositions have determinate 'content' in the sense specified. This is an assumption also implicit in much other work on dispositional theories of meaning. The Sceptic's reply is simple: there is no reason for thinking that this assumption is correct, because there are no facts that could prove (confirm) that it is.

Do Martin and Heil at least have the means to dispel the second criticism, namely that they have failed to give a satisfactory *explanation* of dispositional states' projectivity? A dispositional state acquires determinate content by excluding a potentially infinite number of alternative disposition lines, they say. This exclusion, Martin observes, '… may sometimes be contingent, and sometimes may be necessary. It seems to be necessary in the case of a manifestation in the production or continuance of some *determinate* property that would necessarily exclude the manifestation of another *determinate* property under the same *determinable* obtaining at the same time and place.' (Martin 1993, p. 516). The idea here is this: the manifestation by an object of the property of being-25°-warm necessarily excludes the manifestation, at the same time and place, of the property of being-24°-warm, as well as that of indefinitely many other 'determinate' properties of the same 'determinable' form 'being-*X*°-warm.' This type of necessary connection, incidentally, is at the root of the so-called colour-exclusion problem, which motivated Wittgenstein to abandon his logical atomism of the *Tractatus* period. One should like to think, though, that there could conceivably be ways *other* than this one in which dispositional states are related to their manifestations and the manifestations of other dispositions. For instance, the dispositional state of meaning *plus* by '+' precludes its bearers from replying '5' to the query '67 + 56 = ?' The utterance '5' simply cannot be a manifestation of *that* disposition—if it were, the disposition would not be a disposition to mean *plus*. Therefore, if a bearer of the *plus*-disposition does in actual fact so answer, then we know that, necessarily, some type or other of defeater, for instance another disposition, is active by temporarily blocking or interfering with the *plus*-disposition. Why not assume that the necessity in question is (grounded in) mathematical necessity?

Not so Martin and Heil. In a rather hermetic passage, they explain that 'Dispositionality with its disposition lines directive for a bounded infinity would seem to satisfy the 'and so on' of recursive functions. Knowing a line, one could move from one place (with a specific set of reciprocal disposition partners) to any other place (with a different set of partners, actual or non-actual) along the line. This suggests that recursion is built into nature at the simplest, most basic level. ... dispositionality could ground entailment and mathematical necessity' (Martin and Heil 1997, p. 306). After all, disposition lines can be expected to partially overlap and intersects with other lines. Thus, logical inconsistency could be explained by two different disposition lines intersecting in such a way that the manifestation of one disposition is in conflict with the simultaneous manifestation of the other. 'Conflict' in the realm of dispositions would be the mutual exclusiveness, at a moment in time and space, of a given disposition's manifestation with that of another, as in the temperature example. All this would be going on without us necessarily being aware of it, or having any cognitive access to it: the bearer of a given disposition cannot always recognise that the disposition line of his disposition overlaps and/or conflicts with the disposition line(s) of one of his other dispositions: 'The actual seminality of the disposition here is what grounds a naturalistic account of the objectivity of mathematics and logic and also for the sense of real discovery and failure of discovery [*sic*]. From the self-identity of distinct disposition lines flow the necessities of their overlappings (or points of conflict)' (Martin and Heil 1997, p. 307).

Such a bold foray into largely uncharted territory amounts to 'a new theory of everything', as David Lewis has described it (Lewis 1997). Even more reason to expect particular attention on the part of its authors to what must surely be the centrepiece of this new dispositional realism, namely an analysis of the relationship between a given dispositional state and its disposition line, or between a disposition and its manifestations. Martin (1993) contains a rather remarkable passage on this topic, which we quote in full:

> One should see that the dispositionality *for* a set of manifestations will also *not* be for or even be *prohibitive* against assisting an infinite number of *other* manifestations. This is the basic ontology for setting the *limits* of the infinities of directivenesses and selectivenesses, whether the entities or states are psychological or *non*-psychological, and even whether they are systemic or non-systemic (e.g. elementary particles). This account of the dispositional as

> directive, selective readinesses with the Correlativeness of disposition and manifestation is such that the manifestation 'carries' the richness of the disposition base it is *from* or *of*. It may be only a tip but it is a tip *of* an *iceberg.* It must be remembered that what is being considered is a *present* disposition-base ... that, in a sense, is capable of more than it *could* ever manifest, because on any occasion some forms of manifestation-conditions or reciprocal disposition partners are lacking and may even *exclude* one another. The totality of this infinity of *alternative* manifestations is unachievable, and this is a necessary fact of nature—the *actual* disposition is infinite in its directedness for the manifestations *for* which it is disposed if actual at all are only partial and finite. It is *natural* that so little can carry so much. As a manifestation *of* a particular disposition base, its nature is determined by what it is *from*, namely *that* disposition base with infinite richness of readinesses ...' (Martin 1993, pp. 517-18; all emphasis Martin's)

This sermon-like passage makes a series of strong claims without argument. The main point seems to be that it is a dispositional property's projective essence or nature, which allows it to uniquely determine the set of its manifestations. Substitution of 'projective virtue' for 'projective nature' creates a strong sense of *déjà lu*…

Of course, the rehabilitation of what were once totally unacceptable concepts to the empiricists, namely 'power' and 'capacity'—in other words, the rehabilitation of what the doctor in Molière's *Le malade imaginaire* famously referred to as the "dormitive virtue" of opium—is not restricted to the metaphysics of dispositions and is the work of Martin alone, but it is rather a central plank of the new realism in philosophy of science. It is safe to say, however, that even if explanations, scientific and otherwise, invoking projective, dormitive, and similar virtues turned out to be no legitimate laughing matter after all,[78] the Kripkean Sceptic is likely to remain unmoved if we try to refute him à la Martin by evoking 'selective readinesses' and 'directivenesses'. For the Sceptic's question is—to borrow Martin's metaphor and style of emphasis—*what* iceberg this tip is a tip *of*? The sceptical claim is that there is nothing that will *prove* that *this* tip is the tip of *that* unique iceberg rather than any *other*. In other words, the Sceptic sees no sign of *any* given tip "carrying the richness" of the *whole* and *unique* iceberg beneath it. To put it in a less pictorial manner, the sceptical challenge concerns the alleged infinitary nature of an agent's dispositional states: what facts are there for demonstrating the correctness of the metaphysical picture on

[78] For a succinct presentation of the modern philosophical case for taking dormitive virtues seriously, see Sober, E. (1982). "Dispositions and Subjunctive Conditionals, or, Dormative Virtues Are No Laughing Matter" *Philosophical Review* **91**: 591-596; cf. also Martin, C. B. (1993). "The Need for Ontology: Some Choices" *Philosophy* **68**(266): 505-522, and Mumford, S. (1996b). "Virtus Dormitiva, ha, ha, ha" *Philosopher* **84**(2): 12-15.

offer, of dispositions *qua* dispositions uniquely determining an infinite set of manifestations? In the final analysis, Martin and Heil have little to show for in this respect: they provide an interesting example highlighting the shortcomings of the conditional analysis (the electro-fink argument) coupled with the suggestion that we ought to try something else, and then a metaphysical sermon to the effect that their particular brand of realism about dispositions ought to be regarded as plausible. We must, I suggest, conclude that their case rests on fragile ground. Additional arguments further eroding it will be provided in the third chapter.

2.3 Counterfactual Realism

Martin and Heil's is not the only realist account of dispositions with an explicit eye towards resolving Kripke's paradox. The cornerstone of D.H. Mellor's[79] contribution, just like Martin and Heil's, is the realist claim that dispositional properties are real properties. At any rate, they are just as real as non-dispositional ones, says Mellor. For dispositional properties do not, according to Mellor, entail counterfactual conditionals in a way that is exclusive to them, categorical properties do it too. His well-known example is that of the presumably quintessentially non-dispositional property of being triangular. Mellor holds that X's being triangular entails the counterfactual that if one were to correctly count X's corners, one would obtain the result 3; and so on for every other categorical property (Mellor 1974, p. 171). The moral we are invited to draw from this is that on pain of making *both* dispositional and non-dispositional properties unreal, we must not count their counterfactual entailments against them.

2.3.1 Do Disposition Ascriptions Report 'Conditional Facts'?

Disposition ascriptions are equivalent to statements describing facts, claims Mellor, albeit "conditional facts:"

[79] Cf. Mellor "In Defense of Dispositions", updated in Mellor, D. H. (2000). "The Semantics and Ontology of Dispositions" *Mind* **109**(436): 757-780.

> ... dispositions are real properties in a sense that rules out any account of them as mere potentialities or possibilities. But why should any such account have been thought of in the first place? Dispositional ascription entails statements of (admittedly conditional) fact, not statements of possibility. A fragile glass is one that does break (if dropped), not one that can break. Whether it can break depends inter alia on whether it can be dropped, and its being fragile entails nothing about that. (Mellor 1974, p. 173)

Thus Mellor recommends making the subtle distinction between ascribing a disposition and thereby *entertaining* the physical possibility of the disposition's display, and straightforwardly *asserting* that possibility (ibid.). The former is not tantamount to the latter. Counterfactual, or subjunctive, conditionals do not entail the possibility of their antecedents, and thus are not themselves contradicted by the actual impossibility of their antecedents. Hence we can use them for "entertaining" something contrary to fact: 'Just so we usually resort to subjunctive conditionals only when we think they are counterfactual. Thus dispositions are usually ascribed only when we regard their displays as possible and not actual. The ascription *itself*, though, entails neither of these things.' (Mellor 1974, p. 173; my emphasis). Just so, presumably, we may entertain the possibility of an agent having infinite working memory when we ascribe certain dispositions to her, without however asserting it. We say that if conditions *C* *were* satisfied, the agent *would* (exhaustively) compute the plus function. The subjunctive does not entail the commitment that *C* can actually be satisfied, nor that the agent can compute the plus function—rather, it simply says, according to Mellor, that there is a 'conditional fact' according to which if *C* could be satisfied, the agent could compute the function.

Mellor believes that the epistemology of disposition ascriptions essentially overlaps with questions concerning the criteria for a subjunctive conditional's truth when its antecedent describes a contrary-to-fact situation. It is therefore legitimate, in his eyes, to state the application-conditions of a dispositional predicate in counterfactual terms (see Mellor's analysis of 'fragile', *supra*), and to leave the rather hairy problem of the truth-conditions of the latter to a general solution of the epistemology of counterfactuals. This is an interesting reversal of Goodman's appreciation of the situation, who thought that the problem of conditionals was much harder than that of dispositions, and that tackling dispositions first was possibly a promising way to make progress towards a solution of the problem of conditionals (Goodman 1983, p. 39). In any event, Mellor's ontological stance about dispositions puts him in the vicinity of Martin and Heil 1997 on the issue of 'adding' *vs.* 'quadding', notwithstanding fundamental differences in their view of the correct semantics of disposition ascriptions (Mellor retains the conditional analysis). Thus, although Mellor holds, whereas Martin and Heil reject, the view that disposition ascriptions mean the same as certain counterfactual conditionals, both sides take the realist standpoint that ascriptions of dispositions entail nothing about their possibility of manifestation, and therefore that dispositions can be present in an object under circumstances when their manifestation is physically impossible. Mellor further shares Martin and Heil's view that we need to strictly separate the question of the ontological reality of a disposition from epistemological issues having to do with its manifestations. This leads him to quite similar claims about Don and Van's respective dispositional states.

2.3.2 Reduction Sentences

Mellor 2000, 'an update of his theory of dispositions in the light of recent literature', directly confronts Kripkean scepticism about meaning, and endeavours to satisfy the sceptic's demand for a *fact* about Don that makes it the case that he computes *plus*, by taking recourse to subjunctive conditionals. Mellor holds that of two such conditionals—one of which states the application conditions for 'is disposed to add', whereas the other gives those for 'is disposed to quadd'—only one is true of Don. Crucially, these conditionals must be what Mellor calls (in a tribute to Carnap) 'reduction sentences.' He defines 'reduction sentence' as a sentence that contains the very predicate whose application conditions it is intended to give. For example, the relevant reduction sentence for 'x is fragile' is, according to Mellor,

> 'if x were stressed without ceasing to be fragile, it would break'

According to Mellor, a Carnap-style analysis in terms of reduction sentences is necessary to counter the recent arguments from finkish dispositions and anti-dotes. Mellor acknowledges that reduction sentences introduce circularity, because to understand them we already need to know the meaning of the predicate they define. Though the circle is not vicious, because it '... does not in fact stop us using them to say what dispositional predicates apply to. We can still, for example, remedy the ignorance of those who do not know what to call "fragile" by saying that, by definition, all and only things that remain or become fragile when (relatively lightly and suddenly) stressed will then break.' (Mellor 2000, p. 763).

The relevant reduction sentence for the disposition ascribed to Don by 'means *plus* by '+'' is

> For any two numbers *n* and *m*, if Don were to apply '+' to them *while having this disposition*, Don would get the answer *n+m* (Mellor 2000, p. 764)

This essentially mirrors Carnap's famous reduction sentence for 'x is soluble',[80] with the improvement that Mellor allows the possession by the bearer of the relevant disposition to vary with time. According to Mellor, his reduction sentence for 'means *plus* by '+'' is thus entirely consistent with the circumstance that some numbers are too large for Don to grasp or to add in a finite time, simply because *Don does not always have the disposition to mean* plus. In fact, according to Mellor, it is a general truth about all of us that

> … trying to add [numbers that are too big for us] would cause us to lose this disposition and hence to add them wrongly or not at all. (Mellor 2000, Ibid.).

This is of course quite undeniable: attempting to add two truly enormous numbers will seriously exhaust Don, and if he persists nevertheless, he will eventually die in the process. There is therefore no question that he would add them wrongly or not at all, and that the mere attempt would rob him of this disposition (in the sense of killing him). The question is whether he had it in the first place. Mellor does not explicitly say so, yet what he suggests amounts to the suggestion that Don's disposition to mean *plus* by '+' is *finkish* in Martin's sense—it is an inhibited disposition which, although quite real and present in its bearer, disappears in all circumstances in which its full and exhaustive manifestation is actually called for. Recall that a few queries of the form "58+67=?" alone are *not* such a circumstance, for any finite number of such queries would not constitute a *unique* test condition for the disposition to add—they would also be test conditions for the disposition to *quadd* and infinitely many other dispositions. A full test of adding would be the indefinitely large set of

[80] "If anything x is put into water at any time t, then, if x is soluble in water, x dissolves at the time t, and if x is not soluble in water, it does not." (Carnap, R. (1936-37). "Testability and Meaning" *Philosophy of Science* **3-4**, p. 53; cf. also Section 2.1.1 *supra*),

queries of the form "m+n=?", and attempting to reply to all of these queries will of course make me loose not only my cognitive, but all of my dispositions.

2.3.3 Omniscience (Epistemological Worries II)

Mellor's proposal is unsatisfactory, for epistemological reasons. It seems that we can substitute, *salva veritate*, almost anything for *X* in the sentence-form

> Trying to *X* would cause agent A to lose the disposition for *X*-ing, and hence to *X* wrongly or not at all

Take for instance *omniscience*, which for the purposes of this argument I define as the disposition to reply truly to any query. It would seem that on Mellor's account of dispositions, I can confidently assert that I am omniscient without rendering the corresponding reduction sentence false. For, whenever someone asks me a question that is too difficult for me, I loose my omniscience. In other words, the following claim would be true of me:

> Trying to [*answer any question*] would cause me to loose my disposition to [*answer any question*], and hence to [*answer any question*] wrongly or not at all.

Just like the wire connected to the electro-fink, I regain my omniscience as soon the enquirer goes away, and (in the spirit of this kind of realism about dispositions) I can therefore be said to be omniscient. For, Mellor's reduction sentence for omniscience,

> For any query A, if I were to asked to reply to A *while being omniscient*, I would reply with A's correct answer, B.

would be true of me. In the text, Mellor acknowledges that his reduction sentence has a problem with circularity, and correctly points out that there is a sense in which it is at least not *viciously* circular, because we may actually use the sentence to explain the meaning of the word 'omniscient' to someone who does not know it. The fatal flaw lies elsewhere: without a workable criterion for eliminating unwanted reduction sentences (in other words, without further specification of how to ascertain the obtaining of conditional facts), appeal to such sentences, whether viciously circular or not, is *vacuously true*.

Against this charge, Mellor holds[81] that we need to distinguish two questions here, one ontological, and the other semantic or pragmatic. The ontological question, according to Mellor, is whether the property postulated by the reduction sentence exists (in our case omniscience). Mellor has argued elsewhere that we ought to decide this question by checking whether the relevant property occurs in laws.[82] If it does not, it will, he claims, be an indicator that we need to amend the reduction sentence to stop it being made true by "finkish" properties. The semantic or pragmatic question, on the other hand, is how commonly the reduction sentence is *true*, in other words, how often the relevant dispositional property is "finked". (Recall that to be "finked", for a dispositional property, is to be in circumstances such that whenever a manifestation of the disposition is called for, the disposition ceases to exist; when the circumstances pass, the disposition returns; being disposed to add is a finked property). If the property is finked too often, we will, Mellor says, usually doubt its existence or think that the relevant disposition ascription needs changing. The reply merits three comments: (1) if the answer to the ontological question is decided by appeal to *de facto* laws, many dispositions will be ruled out of existence. There are no extant laws governing courage, for example. Any appeal to *de jure* laws—the as yet to be established laws of a completed science—however, would inherit the problematic nature of the concept of a complete science. Moreover, it would entail that we need to suspend our judgment in most problematical cases until the advent of such a science. (2) The ontological question and the semantic one seem to overlap: the truth or falsity of the relevant reduction sentences will, as Mellor allows, often lead to ontological conclusions. (3) Yet, the fact that in some cases, answers to the semantic or pragmatic question do *not* lead to the corresponding ontological conclusions, is mysterious. Mellor points out that if we suppose, plausibly, that every force applied to a body slightly alters its (Newtonian) mass by knocking a bit from it off, so that every corresponding reduction sentence would turn out false, this would not show that there is no

---

[81] The views and arguments attributed to Mellor in this paragraph are from personal communication.

[82] What properties there are, according to Mellor, can be found out by conjoining all statements of laws of nature, and replacing all the predicates in this conjunction with variables. This then yields a Ramsey sentence which says that 'there are in the world properties that occur in this and that way in laws of nature'. There are no other properties beyond those that occur in laws (cf. Mellor *The Facts of Causation*; and Mellor, D. H. (2001). "Realistic Metaphysics. An interview with D. H. Mellor by Anna-Sofia Maurin and Johannes Persson" *Theoria* **67**: 4-21; accessed on 20/06/2003 at http://www.dar.cam.ac.uk/~dhm11/Theoria.html).

such property nor make its dispositional specification by Newton's laws of motion vacuous. I suppose that on Mellor's view, it would be up to science, again, to tell us which such permanently finked dispositional properties ought to be accepted, and which not (mass, obviously, is accepted). The point of (1) would then apply here as well.

Mellor's account of Don's arithmetical abilities can be classified as but a realist variation of Martin and Heil's. The fundamental idea in both these approaches is the same: Don is the bearer of a disposition which, *in the absence of defeaters*, ensures that whenever he is computing '*m*+*n*', he obtains the correct answer *m*+*n*. Even though defeaters are always and necessarily present (due to our finitude), this does not weaken the relevant disposition's entitlement to be considered real. The account's weakness in the face of the sceptical challenge is equally the same: given the necessary presence of defeaters, we have not been provided with what is urgently required, namely criteria for distinguishing (whether in principle or in practice) ascriptions of acceptable 'finkish' or 'blocked' dispositions, from unacceptable ones. Now, philosophers of a realist persuasion are usually keen to stress that it is important not to 'confuse one's metaphysics with one's epistemology' (Fodor 2001). The warning is well taken, insofar as it can never be a good thing to confuse one thing with another. In our present context it is clear, though, that however great the importance distinguishing epistemological from metaphysical questions about dispositions, not doing your epistemological homework can get you into trouble when you confront a Sceptic. In particular, both Mellor's as well as Martin and Heil's strenuous avoidance of what they see as the traditional empiricist (over)emphasis on how we ascertain a disposition's presence renders their brand of realism fatally incapable of countering Kripke's scepticism. Kripke's Sceptic will ask: if for Don to mean *plus* by '+' is constituted by the obtaining of the conditional fact that

> (A) for any two numbers *n* and *m*, if Don were to apply '+' to them while having the *plus* disposition, Don would get the answer *n*+*m*,

then how are we distinguish the obtaining of *that* fact from the obtaining of

> (B) for any two numbers *n* and *m*, if Don were to apply '+' to them while having the *quus* disposition, Don would get the answer $n \oplus m$?

A realist can of course consistently claim that the two are indistinguishable, and that may be so in principle. After all, a fact's observability or verifiability by us is not a prerequisite for its obtaining. This is the gist of the admonition to not confuse one's epistemology with one's metaphysics. This is, however, not the Sceptic's point. He asks, If you cannot tell me how to distinguish, at least in principle, the obtaining of (A) from the obtaining of (B), why should I *believe* you when you tell me that it is (A), and not (B), that obtains? In particular, how do you justify your claim that (A) obtains?

Mellor correctly points out that the question of the truth-conditions of sentences expressing 'conditional facts' such as (A) and (B) will be solved when that of the truth conditions of subjunctive conditionals in general is solved. But, surely, merely pointing this out and then going on to discuss the metaphysics of dispositions, is misunderstanding the nature of the sceptical challenge? True, the Sceptic makes claims that are ultimately metaphysical in nature. However, the rules of the game, if the realist agrees to play it, require her candidate solutions to demonstrate the existence of facts susceptible to justify corresponding anti-sceptical meaning-claims. A solution hence has to show why we *should believe* that you mean *plus*, given that according to all the available facts you could also mean *quus*. In the absence of instructions on how to establish whether a conditional fact obtains in a given case it is difficult to see how merely postulating (A) will *justify* the claim that Don means *plus* by '+'.

Of course, Mellor as well as Martin and Heil are not to blame for having failed to provide clues on how to empirically verify Don and Van's dispositional/conditional difference—information, say, on how to construct an *experimentum crucis* to pick them apart. We are here not dealing with directly observational truth-conditions, or manifest matters of fact, nor with questions of experimental technique. Moreover, it is part of *any* realist's anti-verificationist platform that existence does not imply verifiability or possible observability, and given verificationism's well-deserved demise, this must be granted. Any preference for endorsing (A) rather than (B), or for ascribing the deep-seated disposition to add rather than the disposition to quad or other spurious candidates, must rather be based on *theoretical considerations*. Their failure to address this point is the most serious shortcoming of the realist solutions discussed so far. The importance of the issue of theory vs. observation in providing a dispositional account of meaning shall become even more salient as we discuss teleological realism.

2.4 Teleological Realism

Ruth Garrett Millikan thinks that Kripke's challenge is an important one. 'The naturalistically inclined philosopher, who … holds intentionality to be an objective feature of our thoughts, owes as solution to the Kripke-Wittgenstein paradox,' [83] she acknowledges, and takes up the challenge. According to her, the lesson to be drawn from Wittgenstein's rule-following argument—and by extension, from Kripke's version of it—is that we need to enrich our conception of intentional action with the notion of a non-represented intention. Millikan's central idea is that

to *mean* to follow rule R = to *have as a purpose* to follow rule R,

and that "having a purpose" is not necessarily an explicitly represented affair on the part of the agent. She introduces a three-fold distinction of the ways in which one can follow a rule:

> (1) merely coinciding with a rule (this is the way in which we conform to "quus" rules and to rules to which we have mere dispositions to conform), (2) purposefully following an explicit or expressed rule, and (3) purposefully conforming to an implicit or unexpressed rule. Way (3) involves having an unexpressed purpose to follow a rule and *succeeding* in this purpose. It is the same as displaying a *competence* in conforming to the unexpressed rule or displaying an *ability* to conform to it. (Millikan 1990, p. 329)

The interesting sense for us is (3) and what makes it different from sense (1), a distinction on which the bulk of Millikan's rejection of Kripke's paradox will be based.

2.4.1 Rule-following, Biological Purposes, and Competence

In order to bring the contrast between merely coinciding with a rule and having an unexpressed purpose to follow it, or a competence, into sharper focus, Millikan uses results from work by Collett and Land on the behaviour and physiology of

[83] Millikan "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox", p. 323. Millikan quotes Loar, B. (1985). "Critical Review of Saul Kripke's *Wittgenstein on Rules and Private Language*" *Noûs* **19**, p. 280.

hoverflies.[84] A male adult hoverfly on the outlook for a mating partner usually hovers in one spot, we are told, in order to keep its flight muscles warm and be able to quickly accelerate and intercept any passing females. Successful interception and subsequent reproduction requires the male to carefully choose its approach trajectory, lest it miss the speedy female. In fact, the fly needs to "calculate" its flight trajectory according to a constant geometrical rule, which takes into account factors such as the male's rate of acceleration, the average speed of females, and the angular velocity of the target's image on the retina of the male. This rule is what Millikan coins the 'proximal hoverfly-rule'. The case is well suited, comments Millikan, precisely because it is so implausible that male hoverflies have an explicit internal representation of the geometrical rule they are conforming to when successfully catching a female. (Just as implausible, we might add, as it is that, say, a child has an explicit internal representation of the physics involved in riding a bicycle). Rather, the fly's perceptual-cognitive mechanisms involved in female-catching have an unexpressed and unrepresented *biological purpose* to conform to the specific "hoverfly-rule" for intercepting females: 'That is, the hoverfly has within him a genetically determined mechanism that historically proliferated in part because it was responsible for producing conformity to the proximal hoverfly rule, hence for getting male and female hoverflies together' (Millikan 1990, p. 331).[85]

The existence of this mechanism could of course explain various *other* dispositions of the hoverfly, such as e.g. the disposition to attract predators by his movements, or the disposition to 'play specific mathematically describable patterns on his retina,' etc. (ibid.). However, what distinguishes these latter dispositions from the former is that they are not liable to explain why the mechanism has survived and reproduced: 'Conformity to the proximal hoverfly rule, on the other hand, has helped to explain the reproductive success of (virtually) every ancestor hoverfly, hence to explain the continued presence of the mechanism in the species.' (ibid.) Incidentally, Millikan calls the relevant hoverfly-rule 'proximal,' because it does not actually specify how the fly should behave with regard to 'distant' objects such as females—the rule only explicitly refers to internal retinal images, not to females. True, by conforming to the proximal rule, a male hoverfly, if successful, also conforms some such 'dis-

---

[84] Collett and Land (1978). "How Hoverflies Compute Interception Courses" *Journal of Comparative Physiology* **125**: 191-204.

[85] Elsewhere, Millikan expresses the same idea by using the term 'proper function', cf. Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Cambridge, Mit Press; also Millikan, R. G. (1989). "Biosemantics" *Journal of Philosophy* **86**: 281-297, p. 152.

tal' rule as 'If you see a female, catch it,' and there is an interesting relationship between proximal and distal hoverfly rules. But the complexities of Millikan's account shall not concern us, for they are immaterial to the explanatory principle at work when Millikan claims that hoverflies follow 'hoverfly-rules' rather than any of their multiple *quus*-like cousins.

Dispositions are plentiful, *competences* are few and far in between. Any given hoverfly has a myriad of dispositions to behaviour, which, however, its organism has no biological purpose to display, says Millikan—such as the disposition to squash when stepped on. What determines the having of a specific biological purpose, or competence, is precisely whether the relevant behaviour, through which the competence is "expressed", can be explained in terms of the evolutionary history of the species.

> This hoverfly displays a competence in conforming to the proximal hoverfly rule when his coinciding with it has a "normal explanation", that is, an explanation that accords with the historical norm. That his behavior coincides with the rule must be explained in the same way, or must fit the same schema, that accounted in the bulk of cases for the historic successes of his ancestors in conforming to the rule. Presumably this normal explanation makes reference to the way the hoverfly's nervous system is put together, how it works, how it is hooked to the retina and muscles, etc. If the hoverfly ends up coinciding with the rule not because his nerves and muscles work in a normal way but only because the wind serendipitously blows him around to face the right direction, he fails to express a competence (Millikan 1990, p. 332)

True, it is always possible to *redescribe* any aspect of the fly's actual behaviour as its conforming to some strange '*qu*overfly-rule,' just as it is possible to redescribe our finite mathematical behaviour in a myriad of quus-like forms. However, conformity to one of these "*qu*overfly-rules" (Millikan's example is a rule exactly like the proximal hoverfly-rule, except that it instructs the fly to sit still if the target's angular velocity happens to be 500°-510° per second) cannot be the unexpressed biological purpose of the fly's physiological mechanisms. The fly cannot have a competence for conformity to *that* rule, because there is no biologically 'normal' explanation available. After all, the disposition to dart off in every female's direction, *except* when the female's image on the male's retina happens to move at a certain speed, does not help to account for the proliferation of the male's ancestors—it does not, at any rate, account for it *as well* as the disposition to dart off in the female's direction *tout court*, and hence the latter is to be preferred in explanation.

In principle, the biological purpose of any organism or part of an organism is always to conform to *those* rules among the millions of rules the organism's actual behaviour happens to be compatible with that *best explain* the past selection of that organism or part of organism. Among all possible descriptions of the organism's behaviour, we must use those which make the most explanatory sense, as it were, from an evolutionary point of view. In other words, our choice of description, and hence of hypothesis, is always the result of an inference to the *best* explanation: we consider true the description which, among all those hypotheses that are compatible with the evidence, is explanatorily superior, according to a previously established standard. Millikan's explicit standard, here, is what she calls "biologically normal explanation". Interestingly, it is immaterial if some, most, or all individuals of the relevant species ever actually conform to these best-explanation-descriptions; we know that real world performance is always flawed, and in this respect "quus-like". Thus, Millikan points out that even if, due to some mechanical engineering constraints, all average male hoverflies were *in principle* unable to pursue those females the images of which happen to move across their retina at exactly 500-510° per second (and thus if all males actually conformed to the proximal *qu*overfly-rule), it would nevertheless be the case that they biologically purpose, and thus express a competence, to follow the hoverfly-rule. Conformity to the hoverfly-rule is, as Millikan says, the biological norm, or 'ideal' (Millikan 1990, p. 335).[86] A biological ideal is not the same thing as an historical average, the latter being affected by (biological and other) accidents.

Millikan concludes: 'This is how purposes inform the rule-following behaviour of the hoverfly, how norms, standards, or ideals apply to his behaviors, hence how the hoverfly comes to display competences or abilities to conform to rules rather than mere dispositions to coincide with them' (Millikan 1990, p. 337). We see how this sort of account would be applied to the rule-following competence of Don and Van. Don, the *plus*-follower, might very well have the disposition to coincide with the *quus*-rule that Van is following. In other words, he does "follow" the *quus*-rule in terms of sense (1) of Millikan's three-fold distinction of rule-following. This does not make him follow the *quus*-rule in sense (3), however, or express the biological com-

[86] In the whole article, Millikan uses the term 'ideal' only twice, in stark disproportion to the importance of the *ideal*–non-*ideal* distinction in her account. Ascriptions of biological purposes, or competences, are of course *idealizations*, a fact that Millikan, along with many other realists about dispositions, never adresses.

petence to do so. True, in the case of humans the complete story is likely to be rather more complicated, because we are more complicated creatures than hoverflies, and the unexpressed biological rules we conform to when doing arithmetic, or buying cinema tickets, etc., are likely to be inordinately more intricate than the hoverfly-interception rule. Moreover, some at least of these rules are not likely to be innate, as in the case of the hoverfly, but acquired. The theory thus needs to make space for learning. But, again, there is no need for us to delve into the details of Millikan's description of the interplay between distal rules, proximal rules, and '*derived* proximal rules' (which she introduces to account for learning)—the basic idea is already prominent. Notwithstanding the complexity of human behaviour,

> … there must be a finite number of proximal and distal "Homo sapiens rules" that we have as biological purposes to follow, and there must be mechanisms to implement these rules built into the basic body and brain of normal persons. … So, unless doing arithmetic results from a total breakdown of the cognitive systems (in which case there may be nothing you purpose when you encounter "plus": how you react to it is accidental under every description) then whatever you mean to do when you encounter "plus", that content has been determined by your experience coupled with evolutionary design. (Millikan 1990, p. 343)

Given that both Don and Van are members of the species *Homo Sapiens Sapiens*, that they are part of the same population and thrive in the same environment exerting the same selection pressure on them, they both share the same evolutionary history. If, as Millikan holds, what Don and Van mean by '+' is determined by that history in conjunction with their own idiosyncratic experience, then if Don and Van mean different things, Don's learning experience must have been different from Van's. If we hold experience constant as well, then both must mean *plus*, lest we have 'total breakdown of the cognitive systems,' i.e. a pathological case. (Which is indeed what would probably cross the mind of anyone confronted with a real-life Van).

Millikan's solution of the paradox, in a nutshell, is this: explicit meaning is accomplished by means of inner representations, which presuppose an underlying and unexpressed biological purpose—namely 'purposing to let the representation guide one in a certain way' (Millikan 1990, p. 343). However, the infinite-regress-of-rules argument cannot take a grip, because no sceptical ambiguity about the meaning of our inner representation of '+' can come up: there is only *one* biologically normal way to be guided by, and hence to use, our representation of '+.' Crucially for our purposes, Millikan acknowledges that this sort of claim makes a few substantial assumptions about explanation:

> …I don't have any particular theory of the nature of explanation up my sleeve. But surely, on any reasonable account, a complexity that can simply be dropped from the explanans without affecting the tightness of the relation of explanans to explanandum is not a functioning part of the explanation. For example, my coat does not keep me warm because it is fur-lined *and red,* nor because it is furlined *in the winter,* but simply because it is fur-lined. (Millikan 1990, p. 334.)

To her credit, Millikan is exceptional among those realists about dispositions who take up Kripke's sceptical challenge by being entirely explicit about her ontological commitments, as well as the obvious link to Goodman-type problems. She continues:

> True, I am making the assumption that the qualifications and additions that convert the proximal hoverfly rule into the proximal quoverfly rule are objectively qualifications and additions rather than simplifications. This assumption rests upon a metaphysical distinction between natural properties and kinds and artificially synthezised grue-like properties and kinds or, what is perhaps the same, depends upon there being a difference between natural law and mere *de facto* regularity. (Millikan 1990, *ibid*.)

Millikan obviously believes, however, that the project of solving Kripke's paradox can be undertaken without defending what she calls 'common-sense ontology', or finding a solution to Goodman's paradox—for neither of these tasks ought to be confused, she warns (Millikan 1990, p. 334). As we have argued throughout Section 1.3, at least the first and the latter task (to say nothing about ontology), although certainly not to be confused, can be linked to the same underlying problem, with the consequence that a solution of one would be unlikely to not be also a solution of the other. Millikan implicitly acknowledges as much in our quote: her position, as we have seen, is that the hoverfly's ancestors managed to proliferate because their behavior coincided with the hoverfly-rule rather than the quoverfly-rule. This claim presupposes that "complexities" can be dropped from the *explanans* without harm to the quality of the explanation, which in turn presupposes that we can draw a metaphysical distinction between grue-like and natural properties. As she points out, this is likely to be tantamount to finding the difference between law-like regularities and grue-like ones, i.e. to solving Goodman's paradox. Millikan thereby in effect presupposes that a solution to Goodman's paradox has/can be found in order to solve

Kripke's paradox. This is rather unfortunate, if our present thesis is correct concerning the interrelatedness of the paradoxes—but perhaps her proposed solution to Kripke's problem also provides us with the means to solve Goodman's? This also seems unlikely, as there are good reasons for thinking that inductive inference and the sort of inference to the best explanation that she employs are closely related modes of inference (see Sec. 2.4.2 *infra*, and Chapter 3).

Millikan purports that her central claims are as *factual* as can be, or at least as factual as the claims of any other science: meaning-facts, on her theory, are essentially psychological facts concerning the individual's learning history and experience, *cum* biological facts about the nature of our cognitive endowments and their biological purpose to follow *Homo Sapiens*-specific rules as they have been shaped through natural selection. To Millikan's mind, these facts are obviously *there,* they are as real as any other sort of fact that biology, psychology, and the other empirical sciences investigate. Indeed, it is not the philosopher's '…task to speculate about the precise form these *Homo sapiens* rules take, or about how the experience of standard training in arithmetic elicits from them the capacity to mean plus. Speculation about the specific forms that our most fundamental cognitive capacities take is the psychologist's job' (Millikan 1990, p. 343). It is not really up to the philosopher, that is, to actually provide Kripke's sceptic with the sort of fact he requests, because the request is at bottom an empirical one, to be satisfied by the empirical sciences. The philosopher can produce no more than, perhaps, a feasibility study—an argument which shows that these facts exist and can, at least in principle, be discovered, and that is exactly what Millikan seems to take herself to have done.

2.4.2 Competences and Deep Dispositions

Contrary to appearances, there are many similarities between Millikan's solution and those of Mellor and Martin and Heil; consequently, the difficulties are equally similar. To start with the obvious: both Millikan as well as Mellor and Martin and Heil believe that what distinguishes Don from Van, from the point of view of the question which rule they are following, or what they mean by '+', are some of their *properties* (or rather facts about their possessing or not possessing these properties). They are all realists about these properties, believing that the question whether an agent exemplifies one of them whereas another does not, is a matter of *actual* difference between them. Beyond this very general agreement about the reality of intentional states, there is of corse potential disagreement as to how to individuate the relevant properties. Martin and Heil appear to believe that "meanings are in the head," namely a part of the complex dispositional array that constitutes the brain of the agent, whereas Millikan is an externalist, suggesting as she does that the only way find out what a given individual agent means is by taking into account biological facts intrinsically external to the agent's organism, e.g. facts about the evolutionary history of the species to which he belongs. Mellor's position on this count is difficult to discern. He claims that the difference between Don and Van, though a difference in their dispositions, is to be analysed as a difference of 'conditional fact', but leaves the question open of how to analyze the truth conditions of subjunctive conditionals. Hence he does not commit himself on the question whether Don and Van's meaning properties are intrinsic to them, or somehow relational.

There is much more than this, however, that links Millikan to Mellor and Martin and Heil. It is precisely the element of her theory that, teleosemanticists would insist, constitutes a radical difference between themselves and "merely" dispositional meaning-theorists. On Millikan's account, which rule an agent is following depends not on actual dispositions, but rather on which *competence* to follow a rule the agent expresses. A competence to do X is, on her story, precisely something very different from a given (set of) disposition(s). Certainly, competences *give rise* to certain dispositions, but which dispositions these are is a factor not only of the competence of the individual, but also of a whole variety of contingent influences. Competences are hence not themselves dispositions, they are not even a special kind of disposition. Rather, they are quite intrinsic, or essential, properties of organisms with which the latter are endowed in virtue of their evolutionary history. In a certain sense, it makes them what they are. For example, Millikan suggests that *any* male hoverfly, in virtue of the very fact that it is a male hoverfly, has 'a biological purpose or competence' to follow the proximal hoverfly rule:

> … The normal hoverfly has a disposition to dart off when it sees a flying bird—and also a disposition to squash when stepped on—but these dispositions do not correspond to biological purposes or to competences. Conversely, male hoverflies that are crippled or blind have no disposition to conform to the proximal hoverfly rule, but still it is one of their biological purposes to do so. As male members of the hoverfly species, conforming is the biological norm, the standard for them. (Millikan 1990, p. 333)

A staunchly empiricist philosopher might be tempted to comment, here, that any view according to which a crippled fly nevertheless possesses an unscathed competence to follow the hoverfly rule amounts to ascribing some sort of *Entelechy* to the fly. However, this would be off-target, insofar as competences as opposed to Entelechies are thought to be well within the scope of empirical, and in particular, biological enquiry. They are in this sense entirely unmysterious and unmetaphysical—just as a native speaker's hypothesized grasp of Universal Grammar is purported to be a fact entirely unmetaphysical. Admittedly, a biological purpose is more deep-seated than any actual disposition an individual may or may not possess, in the sense that it cannot easily be affected by contingencies. As in any traditional teleo-functional theory, Millikan's notion of biological purpose, or competence, performs the role of distinguishing between behaviour that is accidental (such as e.g. the heart's pumping noise), and behaviour performed in conformity with the object's evolutionary *function* (the heart's circulating blood). It is instructive to briefly compare this role with the competence/performance distinction in linguistics we have already referred to above. Like Don's biological competence to follow the plus-rule, Don's competence in English—i.e. his competence to follow the syntactic and semantic rules of English—is thought to be something conceptually rather different from his actual English speaking and understanding performance. Performance in both cases is deemed accidental, and it *underdetermines* competence, in the sense in which we cannot tell merely by examining an individual's behaviour, which competence it is an expression of. In both cases it is claimed that competence can subsist in an agent even when performance is, for contingent reasons, non-existent. Finally, and very importantly for our purposes, both teleosemanticists and generative linguists represent competence as something that, although it *gives rise* to performance, does not do so on its own: performance is the product of competence *plus* various contingent factors, both internal as well as external.

By now we ought to be struck by an obvious parallel between all realist accounts of dispositions. For on Martin and Heil's theory, deep-seated dispositions, too, *give rise to* reciprocal abilities—i.e. finite, more "superficial", dispositions—to do things that are associated with the deep-seated disposition. Just like Millikan, Martin and Heil claim that due to various contingent circumstances, an agent may loose all superficial dispositions connected with his deep-seated disposition, but nevertheless retain the deep-seated disposition. Mellor also mirrors the competence/performance distinction. For according to Mellor, although the conditional

> for any two numbers *n* and *m*, if Don were to apply '+' to them *while having this disposition*, Don would get the answer *n+m*

is true and expresses a (conditional) fact about Don, it is also true of Don that trying to add huge numbers would cause Don to lose this disposition and to add them wrongly or not at all. In other words, Don actually has the disposition to add, but he is incapable of manifesting it. Mellor's underlying idea here is equivalent to the competence-performance distinction. The hoverfly story, for example, could be told entirely in terms of Mellor's 'reduction sentences'. Thus, for any male hoverfly, it is true that

> For any image of a certain size moving across the hoverfly's retina with a certain speed, if the hoverfly were to follow the hoverfly rule *while having this disposition*, it would dart off in the direction and with the acceleration indicated by the hoverfly rule

However, it is *also* a general truth that

> Trying to effect indefinitely many starts would cause the hoverfly to lose this disposition, and follow the hoverfly-rule wrongly or not at all.

Just like adding indefinitely many pairs of numbers in the case of *plus* vs. *quus*, effecting indefinitely many starts is what is required to behaviourally differentiate following the hoverfly-rule from following the quoverfly-rule.

The linguist's story about competence and performance could also be told in Mellor's conditionals: any English speaker actually has the competence to understand any English sentence, but he or she is incapable of (fully) manifesting it. In other words,

> For any spoken English sentence, if an English speaker were to hear it *while having the competence to speak English,* she would parse it correctly

However, it is also true that

> Trying to parse sentences that are too long for her, would make the English speaker loose her competence

She would die in the process, and a corpse, arguably, has no linguistic competence. Finally, it seems that both the hoverfly-scenario and the linguist's distinction could also be represented in Martin and Heil's terminology of hierarchical inhibited and uninhibited dispositions, an exercise we shall forego.

What is going on here? The upshot of all this seems to be the following: realists about dispositions *qua* realists advance their respective arguments based on the same sort of distinction. Terminology may vary, as may details. Schematically, however, the idea remains throughout as follows:

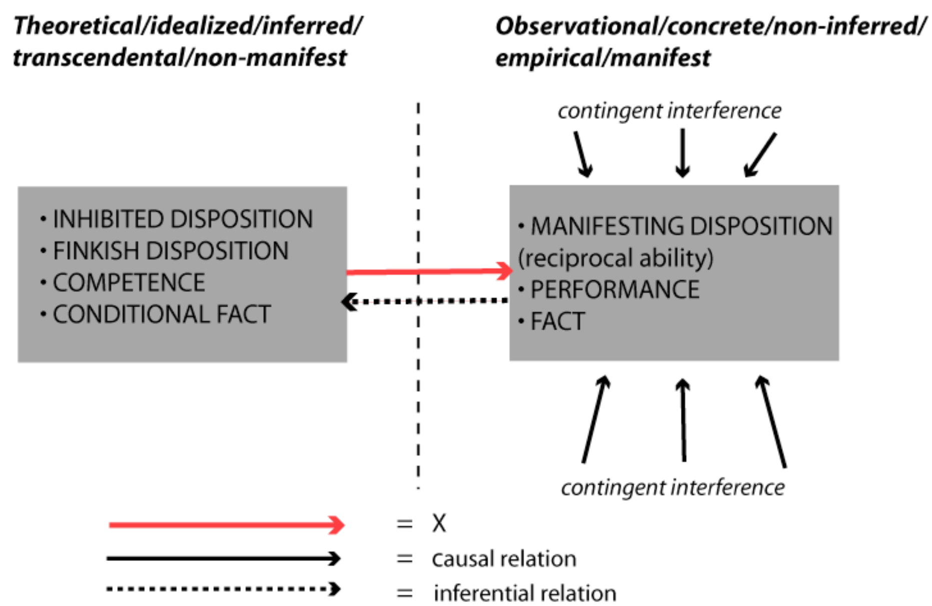

Fig. 7

The multiple headers in this figure are of course not intended to imply synonymy, or to assimilate in any comparable way the conjoined terms. It is obvious that inferred entities are not necessarily transcendental, and that not everything that is not idealized, or concrete, is thereby empirical. Similarly, I take it that the manifest/non-manifest distinction is not co-extensional with the theoretical/observational distinction, and that the latter is problematic in the first place (cf. Kukla 1996; Bogen and Woodward 1988). The terminological abundance here is rather intended to further an ecumenical cause. Thus, the philosopher unfavourably disposed towards the Kantian dichotomy of the transcendental vs. the empirical might accept that of the theoretical vs. the observational, and theorists who do not believe in the latter might accept that of the idealized vs. the non-idealized. The choice of dichotomies is irrelevant to the purposes of the present distinction, insofar as whichever type of dichotomy is preferred, the entities referred to in the boxes will be distributed as depicted.

There is, in other words, an obvious and important difference between the two groups, which can be variously described. It is this difference between them, and their relation to each other, which concerns us. Now, Chomskyan linguists as well as teleosemanticists would vehemently reject the suggestion that their respective notion of competence is nothing but a *special* sort of dispositional concept, namely an idealized one that abstracts away from contingent factors. For to get away from the concept of disposition—in particular, from the empiricist construal of it and its concomitant links to behaviourism—was, within their respective fields, one of the main theoretical motivations of these theories. However, whether competences are idealized dispositions, deep dispositions, or finkish dispositions[87]—whether they are best analysed in terms of conditional facts, or whether they are entities or theoretical constructs of an entirely *sui generis* nature having nothing in common with dispositions—does not matter from our present point of view, if the theoretical role they fulfil within a given account is acknowledged to be the same as the role of inhibited or idealized dispositions. And it seems undeniable that all the concepts on the left side of the table are designed to allow a distinction between actual behaviour that has been affected by contingent causal factors, and non-accidental behaviour that would be an essential expression of whatever the relevant account wishes to be realist about, i.e. deep dispositions, competences, conditional facts, etc. All concepts on the left apply to entities in a *causally isolated* system, whereas all on the right do not. Moreover, all concepts on the left are deployed as a result of an inference to the best explanation (Millikan is the most explicit about this). From this vantage point, the sceptic's reply to each of the proposed solutions—namely that the ascription to me of a particular deep disposition, idealized disposition, competence, or the obtaining of a specific conditional fact about me, respectively, cannot be accounted for by facts about my manifest, uninhibited dispositions, performance, or actual facts about me—represents *the same kind of contention throughout*. Dealing with it head on requires tackling the problem of describing how we project from the factual basis for our disposition-ascriptions to these dispositions themselves, and to identify the legitimate constraints, if any, that govern this process. All the above ascriptions are instances of the same

[87] Jerry Fodor, for one, seems to take it for granted that a theory of idealized dispositions is equivalent to one deploying some sort of competence/performance distinction (Fodor "A Theory of Content II", p. 94 and *infra*). For an argument that explanations in cognitive psychology invoking the concept of 'competence' inherently contain idealizations (and are therefore problematic), see Franks, B. (1995). "On Explanation in the Cognitive Sciences: Competence, Idealization, and the Failure of the Classical Cascade" *British Journal for the Philosophy of Science*: 46(4) 475-502 and Franks, B. (1999). "Idealizations, Competence and Explanation: A Response to Patterson" *British Journal for the Philosophy of Science*: 50(4) 735-746.

sort of inference well within the scope of the Goodman-Kripke paradox—or so I will argue in the last Chapter. First, however, we shall take a look at another prominent dispositionalist reply to Kripke, one that has the double merit of clearly displaying the thoroughly theoretical nature of the sought-after fact of the matter and of clearly establishing the link, already noted by Kripke, to *ceteris paribus*-claims and laws.

2.5. Nomological Realism

In the course of expounding his well-known causal theory of mental content, Jerry Fodor presses the view that causal laws, rather than being relations between individuals, are relations between properties (Fodor 1990, pp. 93-94). A consequence of this view is a methodological point "about which he feels strongly," namely that philosophical theories in terms of nomic relations between properties do not need to be further analyzable in terms of counterfactual relations among individuals. To illustrate, Fodor discusses Kripke 1982 as an example of how a philosopher can be lead into error by implicitly assuming that a causal law statement needs to be analysable into counterfactual truths. Kripke takes the dispositional explanation to run roughly as follows: 'It gives a criterion that will tell me what number theoretic function φ I mean by a binary function symbol '*f*', namely: The referent φ of '*f*' is that unique binary function φ such that I am disposed, *ceteris paribus,* if queried about '*f*(*m*, *n*)', where '*m*' and '*n*' are numerals denoting particular numbers *m* and *n*, to reply '*p*', where '*p*' is a numeral denoting φ*m*, *n*)' (Kripke 1982, pp. 26-27). This sort of account amounts to 'a dispositional story backed by an appeal to the performance/competence distinction,' comments Fodor 1990, p. 94, and he has no principled qualms with it as stated. (It would seem that Fodor endorses our schema of Section 2.4.2, at least insofar as the competence/performance and disposition/manifestation distinctions are concerned.)

2.5.1 Dispositions, *ceteris paribus*

Kripke's discussion makes it plain that he conceives my disposition, *ceteris paribus*, (my CP-disposition, for short) to do B if A to be a matter of the obtaining of certain counterfactual truths about me prefaced with some CP-operator: CP (if A then B). Thus, for me to be the bearer of the CP-disposition to add is for it to be the case that CP (if queried "$m + n = ?$", I would respond with the sum of the numbers *m* and *n*), where 'sum' denotes the value for arguments *m*, *n*, of the binary function addition. Kripke rejects this analysis not on the grounds that dispositions ought not be analysed in terms of counterfactuals—in fact he seems to assume throughout his exposition that this poses no significant problems—but rather because we have no satisfactory analysis of the 'CP' operator:

> … how should we flesh out the ceteris paribus clause? Perhaps something like: if my brain had been stuffed with sufficient extra matter to grasp large enough numbers, and if it were given enough capacity to perform such a large addition, and if my life (in a healthy state) were prolonged enough, then given an addition problem involving large numbers, m and n, I would respond with their sum, and not with the result according to some quus-like rule. But how can we have any confidence of this? … Surely such speculation should be left to science fiction writers and futurologists. We have no idea what the results of such experiments would be. (Kripke 1982, p. 27)

If it is indeed the case for all numbers *m*, *n,* that all other things being equal, I could correctly compute their sum, then, given that it is not actually the case that I can add extremely large numbers, the truth of that *ceteris paribus*-claim requires that there is something about my present circumstances, which makes it the case that other things are not presently equal. Generally, a believer in the truth of any *ceteris paribus*-claim is invited to assume that if what the claim says is not actually true at the moment of utterance, then this must be because the conditions at that moment are such that other things are not "equal." The claim that *ceteris paribus*, A, has thus the form $C_A \rightarrow A$, where 'C_A' refers to a description of the conditions under which everything else is "equal" with respect to the question of the truth of 'A'. Whether $C_A \rightarrow A$ can be interpreted simply as $C_A \supset A$ shall be, in a sense, the focus of the rest of this chapter, for the interesting thing is of course what happens to the truth-value of this claim if C_A is never true, e.g. because it is impossible. What does it mean for a set of circumstances to be "equal" with respect the truth of a given claim A? It cannot mean a *sufficient condition* for the truth of A, for then all claims $C_A \rightarrow A$ would be trivially true. Intuitively, one might think that when we claim '*ceteris paribus*, A' it is actually the case that A and we hold that A will be the case whenever a specific set of conditions C_A are "equal" to actual conditions. However, not all CP-claims are like that.

'C_A' may sometimes be a set of quite ordinary conditions, and sometimes quite extraordinary ones. Take the claim that, everything else being equal, Paul goes to the Pub every day at 7 pm. This is to say that he *usually* goes to the Pub, unless something happens that makes his day an unusual one. Thus, he does not go if Manchester United plays Arsenal on that night, or if the weather is bad, or if he doesn't feel well, or if it is Christmas, etc. This case is sufficiently similar to our intuitive construal $C_A \rightarrow A$, in so far as it is indeed actually the case that Paul goes to the Pub every day at 7 pm, with a few exceptions that are easily accounted for by showing that C_A failed to be the case. On the other hand, we also make claims such as, *ceteris paribus*, two massive bodies b^1 and b^2 will attract each other with the force $F = \frac{m_1 m_2}{d^2}$.[88] Unlike in Paul's case, 'C_A' here is a very *unusual* set of conditions, namely a universe devoid of other bodies, and of other forces (e.g. electrical forces). This means that the claim is never actually true, and it would appear that we must interpret the CP-operator as saying that A will be the case whenever actual conditions are equal to those very special conditions (which *never* happens)—in other words, the converse of what CP means in Paul's case.

Clearly, Kripke's worry about the CP-clause is quite legitimate. We know as a matter of physics and the life sciences that real conditions in this world never even remotely resemble those conditions (if any) under which the relevant conditional about my adding abilities would come out true. Moreover, the attempt to specify *which* sort of circumstances would be just right does appear as little more than specious possible-worlds-speculation. Detractors claim that, generally, the use of the CP-clause is a deliberate fudge, covering up white spots in our knowledge. Kripke's criticism that (a) the relevant *ceteris paribus* claims are true only under quite fantastic counterfactual conditions, (b) we are unable to give even a rough description of these conditions, is therefore well motivated, to say the least. The lesson he draws is that we should take analyses of what it is to mean *plus* by '+' in terms of CP-dispositions for what they ostensibly are, namely science fiction.

---

[88] At least it *seems* that when we assert the universal law of gravitation, we are hedging it with a CP-clause. For more ample discussion of the alleged *ceteris paribus*-character of even our most fundamental laws of nature, see Cartwright, N. (1983). *How the Laws of Physics Lie*, Oxford, Clarendon Press (*pro*); Earman, J. and J. Roberts (1999). ""Ceteris Paribus", There Is No Problem of Provisos" *Synthese* **118**(3): 439-478 (*contra*). For further discussion whether ideal-condition claims, such as the one presumably contained in '$F=m_1m_2/d^2$', are identical with, or a subset of, ceteris-paribus claims, see Section 2.5.1, especially footnote n° 94.

Fodor counters that this is inconsistent with established scientific methodology (Fodor 1990, pp. 94-95). We have many important and rather well entrenched laws whose antecedents refer to ideal initial conditions that cannot obtain. We are, he says, nevertheless confident in the truth of those laws, which assert that *under those ideal conditions*, the consequent follows the antecedent. Fodor's case in point is a traditional philosopher's favourite, the Ideal Gas Law. The law states that the pressure (P) and volume (V) of a gas in a container varies with the number of gas molecules (n) in the container, multiplied by the 'universal gas constant' (R) and temperature (T): $PV = nRT$. This equation will only be literally true if collisions between individual gas molecules are perfectly elastic, i.e. do not decrease their momentum. Thus, it is usually stipulated that the law applies only to ideal gases such that (1) the volume of the molecules themselves is "much smaller" than the volume of the container in which they are held, and (2) the range of the electrical force between molecules is "much smaller" than the average distance between molecules.[89] In other words, the law applies to gas molecules that are essentially considered to be point-masses, and to not interact with each other except through direct collision. But, as Fodor quite rightly points out, 'God only knows what would happen if molecules and containers actually met the conditions specified by the ideal gas laws … for all I know, if any of these things were true, the world would come to an end. After all, the satisfaction of these conditions is, presumably, physically impossible and who knows what would happen in physically impossible worlds?' (Fodor 1990, p. 94). We cannot fully specify the conditions C for which the law holds, for God only knows the all the features of a world instantiating C—but so what? In the sciences we routinely use idealized laws with some success, even without being in God's enviable epistemological position. CP-claims that only come out true under extraordinary circum-

[89] Cf. Orear, J. (1982). *Physik*, München, Carl Hanser, p. 239. It is also sometimes stipulated that an ideal gas simply is any gas that obeys PV = nRT, which would make the claim that ideal gases conform to PV = nRT analytic... In any event, for PV = nRT to hold *strictly*, further conditions need to obtain which physics textbooks do not always mention: for instance, containers need to be absolutely impermeable, and the collisions between the molecules and the walls of the container need also to be elastic. It is in general quite difficult if not impossible to non-trivially and exhaustively specify the conditions under which ideal laws apply strictly. More specific laws, such as Van der Waal's equation, attempt to take into consideration facts about real gases, i.e. that molecules are not point masses and do attract each other, resulting in imperfectly elastic collisions that reduce their kinetic energy and have an effect on pressure. Nevertheless, even Van der Waal's equation expresses a regularity that is merely *ceteris paribus* (cf. Pietroski, P. and G. Rey (1995). "When Other Things Aren't Equal: Saving "Ceteris Paribus" Laws from Vacuity" *British Journal for the Philosophy of Science* **46**(1): 81-110, at p. 97; and *infra*).

stances are not to be rejected on these grounds alone, for this would throw out the baby with the bath, and rule out *most* scientific laws.

Idealization has been a ubiquitous part of scientific inquiry ever since Galileo, and it poses well-known philosophical puzzles. What justifies the introduction of idealizations in science? How could laws that are defined over conditions that never obtain be instrumental to our explanations and predictions of the behaviour of *actual* objects? Pre-theoretically at least, we take established laws to be true statements about nature "as it really is," even when these laws make reference to heavily idealized and abstract objects, such as point-masses, frictionless planes, free markets, ideal speakers, and so on. The fact that the actual, or experienced, world contains no such things, and is *known* to contain no such things by necessity, ought to be puzzling. How could idealized laws be (empirically) confirmed by a reality that is not at all like they purport it to be? If we take the laws of nature to be true statements about nature as it is, does that mean that those laws that do *not* recur to idealizations and merely register correlations between observed events (i.e. experimental or "phenomenological" laws), describe a reality *less* profound, i.e. nature as it appears, not as it really is? How are ideal laws and experimental laws related?

Although he suggests that it is ultimately Kripke's neglect of the important role in science of laws defined over idealized conditions that makes him worry needlessly about how to 'cash out' a CP-law in terms of its counterfactual implications, Fodor 1990 does not actually engage in a discussion of that role, nor does he attempt to answer any of the just mentioned questions. All we are told is that it is *not* necessary, for the Ideal Gas Law to enjoy a good scientific reputation, that '… we know anything like all of what would happen if there really were ideal gases. All that's required is that we know (e.g.) that if there were ideal gases, then, ceteris paribus, their volume would vary inversely with the pressure upon them. And *that* counterfactual *the theory itself tells us is true*.' (Fodor 1990, p. 95; emphasis Fodor's). In Fodor's eyes, scientific practice shows precisely that, *contra* Kripke, we *can* accept generalizations defined over idealized conditions even if we are unable to tell which contingent consequences would follow if these conditions were to obtain. For, knowing that a certain law-like proposition is true does not require knowing which counterfactuals, if any, are "supported" by the law.[90] What Fodor is aiming at, in particular, is the spe-

[90] So-called 'counterfactual support' (i.e. the entailment of counterfactual conditionals) and instance confirmation were traditionally considered necessary, if not sufficient, conditions on any law-like generalization. However, it seems that this would be asking too much of CP-laws. For counterfactual support and instance confirmability are not reliable symptoms of CP-lawlikeness, as Peter Lipton points out: "'All Fs are G, cp' may be a law yet not entail that if something had been an F it would have been

cial case of psychological laws. Psychological laws—among which our hypothesis about the arithmetical abilities of average human agents—are neither better off nor worse off in this respect than other laws and theories, Fodor holds. It constitutes no argument against the validity of the latter hypothesis, according to him, that we do not know what would happen if our working memory really were indefinitely large. The only counterfactual we need to know is again "the one the theory itself tells us is true," i.e. 'If we did have unbounded memory, then, ceteris paribus, we would be able to compute the value of m+n for arbitrary m and n.' (Fodor 1990, p. 95).

Having thus suggested that Kripke's scruples with regard to CP-disposition ascriptions are methodologically unwarranted, Fodor displays no qualms in stating sufficient conditions for a speaker's meaning *plus* by '+' in terms of the speaker's possession of a certain CP-disposition with regard to '+':

> … it is arguably a sufficient condition for a speaker's meaning plus by "+" that, ceteris paribus, he takes "m+n" to designate the sum of m and n; a sufficient condition for a speaker's meaning and by "and" that, ceteris paribus, he takes "P and Q" to be true iff he takes "P" to be true and "Q" to be true; and so forth. (Relations like "taking to express", "taking to be true"—which, on this construal, hold between symbol users and symbols they use—would have to receive a causal/dispositional reconstruction if circularity is to be avoided. (Fodor 1990, p. 111)[91]

Just like Martin and Heil 1997, Fodor therefore holds that what distinguishes agent Don, who means *plus*, from agent Van, who means *quus*, are their differing *dispositional states*—Van, by hypothesis, is differently disposed towards '+' than Don. Of course, unlike Martin and Heil, Fodor has a rich and developed theory on offer about what the possession of mental dispositions actually amounts to in terms of the higher-order structure of the brain (a well-known story about modules, belief boxes, functional roles, etc.)—but this matters little from our present point of view. Both ac-

a G, nor will observed Fs that are G always provide reason to believe that the next F will be a G as well, since we may have no reason to believe that all things will be equal, the next time." (Lipton, P. (1999). "All Else Being Equal" *Philosophy* **74**(288): 155-168, at p. 157).

[91] I should note that Fodor explicitly endorses a dispositional account of meaning only for 'logical vocabulary', a category in which he includes mathematical terms. His theory of the meaning of non-logical terms is a well-known *causal* one, based on the idea that 'cow' means the property of being a cow because the property of being a cow is nomologically connected to the property of being a cause of tokens of corresponding "cow"-symbols in my mind (cf. Fodor *Psychosemantics*; for cricitism see, for example, Loewer, B. and G. Rey (1991). *Meaning in Mind: Fodor and his Critics*, Cambridge, Blackwell). Evidently, it would have been quite hopeless to look for laws relating the property of being the *plus*-function with the property of being a cause of tokens of '+'-symbols in my mind, as mathematical properties are not usually thought to enter into causal relations.

counts take a robustly realist stance about dispositions and construe our meaning one thing rather than another by '+' as a matter of our disposition with regard to '+.' Although Fodor's theory of mind is certainly more illuminating and explanatorily superior to Martin and Heil's laconic characterization of the mind as a 'dispositional array', it is not immediately obvious that Fodor's stance *vis-á-vis* the plus *vs.* quus issue represents any advance over Martin and Heil from the point of view of its resistance to sceptical doubt.[92]

Fodor, contrary to Martin and Heil, admits laws of nature into his metaphysics, in particular psychological laws, and believes that common-sense psychological explanation is "at least sometimes" explanation by law-subsumption. In particular, the following psychological law, defined over idealized conditions, may well be true in his eyes:

> ∀(x) [If x had unbounded memory → CP(x would be able to compute the value of m+n for arbitrary m and n)],

where x ranges over an appropriate sub-set of healthy adult humans having undergone a specific sort of arithmetical training. If we take "to be able to do X" as elliptical for "to have the disposition to do X", then this law warrants a corresponding disposition ascription to these cognitive agents, a disposition that would manifest itself under the ideal conditions specified in the antecedent. Importantly for our purposes, it is the *ceteris paribus* clause that takes care of the putative difference between Don and Van: Van is a human agent who has undergone the same arithmetical training as Don, and is thus ought to be governed by the same law as Don—nevertheless, he would not compute m+n, but m⊕n, if he had unbounded memory. Hence it must be that in *his* case not all things are equal. Thus Fodor discharges the burden of showing why certain disposition ascriptions are obviously spurious whereas others are not, on an analysis of the CP-clause. This makes the degree to which Fodor's theory succeeds in providing a satisfactory theory of CP-clauses (in particular of CP-disposition ascriptions) the critical factor for determining if it constitutes a successful dispositionalist reply to the meaning sceptic.

Fodor 1990 provides us with little relevant detail to decide this question. In fairness, neither CP-laws, nor dispositions—nor idealization for that matter—are his

---

[92] We shall return to the issue of functionalist theories of dispostions in Section 3.2.1.

primary concern there. We need to turn to his reply to an article by Stephen Schiffer[93] in which the latter provocatively claims that there are *no* CP-laws in psychology. Fodor 1991 promises to address our epistemological worries and tell us more about the truth conditions of psychological CP-laws. The problem, to recapitulate, is to show why and how I can be confident of a lawlike statement such as

> I mean ϕ by 'f' ≡ I am disposed, *ceteris paribus,* if queried about '$f(m, n)$', to reply 'p', where 'p' is a numeral denoting $\phi(m, n)$,

although I cannot be confident, and according to Fodor should not *expect*, that the truth of this statement necessarily entails or supports any particular counterfactual conditional, such as, say,

> If queried "8569214 +564879 = ?," I would respond with '9134093'

In the absence of counterfactuals entailed or supported, it is unclear what the confirming instances, if any, are of a law of the form $\forall(x)\ (F(x)\rightarrow \text{CP } G(x))$. (Where '$\rightarrow$' represents a nomological connector appropriate to CP-laws). Suppose that F describes a physically impossible state or event type, and that it is actually the case that $F^*(a)\rightarrow G(a)$. What does the relation of F* to F need to be for $F^*(a)\rightarrow G(a)$ to count as a confirming instance of $\forall(x)\ (F(x)\rightarrow \text{CP } G(x))$? And if there are no such instances, what reasons do we have to believe in the law?

We need to carefully distinguish the problem of CP-clauses from the problem of how to empirically test laws defined over impossible (idealized) conditions, however. The specific difficulty created by CP-clauses is not that they threaten to render empirically unverifiable all statements they are attached to. For if F is an impossible property or event, then '$\forall(x)\ (F(x)\rightarrow G(x))$' will pose this problem even without any CP-operator. Rather, CP-clauses threaten to make laws, and indeed any proposition in which they occur, blatantly *vacuous.* Take again the claim that, everything else being equal, Paul goes to the Pub every day at 7 pm. Obviously, we cannot complete the list of possible events interfering with Paul's going to the Pub, i.e. we cannot fully specify the set of possible conditions which would make it the case that not everything is equal (for Paul in this respect). This problem in itself has nothing to do with idealization. Nevertheless, it is easy to conflate the two issues, because many CP-claims are

[93] Schiffer, S. (1991). "*Ceteris Paribus* Laws" *Mind* **100**: 1-17; Fodor, J. A. (1991). "You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanation" *Mind*: 19-34.

predicated on implicit idealizations—namely when the *de facto* prevailing conditions are "permanent interferers" whose presence needs to be subtracted in thought, and the relevant claim is true only under those rare or physically impossible conditions when the interferers are absent. For example, the unavoidable presence of other bodies is, as we have seen, a permanent interfering condition with respect to the law $F = \frac{m_1 m_2}{d^2}$. We might say, in terms of a useful distinction drawn by Geoffrey Joseph between *ceteris paribus* and *ceteris absentibus* clauses (Joseph 1980), that those CP-clauses that contain implicit idealizations are in fact CA-clauses, i.e. clauses that claim that if certain causal factors were *absent*, although they are in fact permanently present, then if A then B. In terms of our analysis of CP(A) as having the form $C_A \rightarrow A$, the claim amounts to an idealization when C_A is physically impossible.[94]

In order to secure a legitimate role for CP-clauses, we need to show, in Nancy Cartwright's words, that "all other things being equal" does not simply mean "all things being right" (Cartwright 1983, p. 45). Surely, if all things are right for his going to the Pub, then Paul goes to the Pub. Similarly, if all things were right for me, then if queried "8569214 +564879 = ?," I would respond with '9134093', and so on for even larger numbers. In order to make CP-laws respectable, we thus need to show that even though there may always be quite extraordinary conditions under which

[94] I thus sympathize to a certain extent with Earman and Roberts ""Ceteris Paribus", There Is No Problem of Provisos", who hold that idealized laws ought not be understood as *ceteris paribus* laws, because it seems that contrary to the case of the ideal conditions stipulated in the antecedent of an idealized law, the *ceteris paribus* conditions of a hedged law must actually obtain at least sometimes. In my view, this observation correctly represents our pre-theoretical understanding of the phrase "everything else being equal" for cases such as Paul's regularly going to the Pub. However, both CP(A) and CA(A) claims are of the form $C_A \rightarrow A$, and the difference, if any, between them is unclear: it is a well-known difficulty in the metaphysics of facts to say what distinguishes "positive" facts, e.g. the fact that a given set of conditions C is *present*, and "negative" facts, e.g. that C is *absent*. Is not the negative fact of the absence of C tantamount to another positive fact, namely the presence of $\neg$C? It would seem that every positive fact is a negative one under some description, and every negative fact a positive one under another. It is precisely this sort of problem that prompts Pietroski and Rey to define their notion of 'interferer' as *either* a 'positive' factor *or* the 'absence' of such a factor—for, as they put it, they would not even dream of attempting to define the difference between the two (Pietroski and Rey "When Other Things Aren't Equal"). Consequently, no hard and fast line distinguishing *ceteris paribus* from *ceteris absentibus*-clauses can be drawn. This observation, in turn, suggests that no hard and fast line between CP-laws and idealized laws can be drawn: if many CP-laws are in fact *ceteris absentibus* laws, and if the latter involve the (physically impossible) subtraction of causal factors and processes that are in fact permanently present, then the set of idealized laws would seem to be a subset of the set of CP-laws. Earman and Roberts provide a second reason for keeping idealized and CP-laws apart, namely that any attempt at explicitly stating the *ceteris paribus* conditions is doomed to fail, whereas the ideal conditions in the antecedent of an idealized law often appear precisely and completely statable. I shall attempt to throw doubt on this *infra*. (cf. Liu "Laws and Models", who rejects Earman and Roberts first reason, but accepts the second.)

everything would indeed be "right," and, for instance, a turtle would be able to outrun a man, we are not thereby justified in asserting that 'CP (turtles run faster than humans)' is a law, or even a plausible law-candidate urgently calling for empirical confirmation. Generally, if for a given law CP(A→B) all conditions except those under which B follows A are designated as cases where not everything is "equal", then the CP-clause would simply mean 'A→B, *unless conditions are unfavourable*', making CP(A→B) true *a priori*. As Fodor puts it, 'It'll fly, ceteris paribus' ought not to mean the same as 'It'll fly, unless it doesn't' (Fodor 1991, p. 22).

Now, Fodor thinks that any philosopher who has succeeded in formulating general truth conditions for CP-laws "is probably in want of a long rest," and he accordingly limits his ambition to defining the difference between specifically *psychological* CP-laws and other laws, thus taking the notion of lawhood itself for granted (Fodor 1991, p. 22).[95] Assuming that A is a propositional attitude in virtue of which an agent satisfies the antecedent of 'CP (A→ B)', and that B is a type of behaviour in virtue of which he or she satisfies the consequent, Fodor defines as follows:

> Let $A(R_i)$ be an event type in which A is realized by R. Let C be an arbitrary event type. Then C is a COMPLETER relative to a realization of A by R_i iff
>
> $A(R_i)$ & C is (strictly) sufficient for B
> It is not the case that $A(R_i)$ alone is sufficient for B
> It is not the case that C alone is sufficient for B. (Fodor 1991, p. 23)

Applied to our case, the relevant completer for Don's propositional attitude of 'meaning *plus* by '+''would be a specification of conditions C_D, such that, when C_D obtains, and Don is in mental state A realised by neuro-physiological state R_D, then Don is able to add correctly any number *m* and *n*: $A(R_D)$ & C_D → B. The vast majority of authors writing on the topic of *ceteris paribus* conditions are agreed that most CP-law candidates are not susceptible of completion in this way. We will not be able to find, for most putative psychological laws of the form CP(A→B), a condition C for all states A(R) such that instantiation of A(R) and C is strictly sufficient for B. In particular, it does not seem very likely that we will be able to find C_D for human agent Don and his typical mental states. This leads Schiffer to remark that sentences of the form CP(A→B) really have the form CP(A & … → B) and hence fail to express a

[95] However, he expresses the hope that an account of CP-laws ranging over psychological states would 'provide some insight into the way that cp laws [in general] work' (Ibid.).

proposition, let alone a lawlike proposition, and therefore have *no* truth conditions at all (Schiffer 1991, p. 2, and *passim*).

Fodor's proposed solution is as follows: a particular realizing state $R_D(A)$, which realises the attitude A of 'meaning *plus* by "+"' in Don, is either such that in conjunction with conditions C_D it is nomologically sufficient for B, or it is not. If it is, then CP (A→B) is true of Don. But even if C_D happens to be not instantiated, then CP (A→B) can *still* be true of Don. After all, it is not nomologically necessary that the relevant completing conditions C are tokened whenever A is realised by a particular R.[96] CP (A→B) could be a true CP-law about Don even though Don does not instantiate conditions C_D. In this case Don would constitute what Fodor refers to as "a mere exception" to CP (A→B). (Fodor 1991, p. 24). The exception would be "mere" instead of "absolute", because notwithstanding the fact that the relevant realizer R_D for Don's attitude A happens not to be completed in the right way, we may still rightly consider Don to have A. We must allow mere exceptions to CP-laws, for having such exceptions is what makes laws *ceteris paribus* in the first place—allowing them is the whole point of a CP-law, says Fodor. What distinguishes CP-laws with mere exceptions from those that encounter "absolute" exceptions is that we could, at least in principle, strip off their CP-operator by enriching the antecedent with a (finite) list of conditions such that the enriched antecedent together with the laws of nature is strictly sufficient for the consequent.

Much more problematic, and philosophically interesting, are *absolute* exceptions to the truth of a CP-generalization, i.e. those cases when "some or all" realizer states of A cannot be completed, not even in principle. Far from denying them the status of law-hood altogether, as Schiffer and other authors would have it, Fodor proposes a criterion for defining the subset of those absolute exceptions that we should nevertheless be prepared to tolerate. He calls the union of the laws containing A in their antecedents the 'network' of laws for A, and stipulates that if 'CP (A → B)' is one of the laws belonging to the network for A, this law may be considered true if either of the following conditions is satisfied

i. every realizer of A has a completer (that is, there are no absolute exceptions to A→B)
ii. if R_i realizations of A are absolute exceptions to A → B, then there must be many *other* laws in the network for which R_i has completers. (cf. Fodor 1991, p. 27.)

96 This is the essential realist move, equivalent to Martin and Heil's claim that it is not necessary for the instantiation of a disposition that conditions be such that it can actually manifest.

In Fodor 1990 he had maintained against Kripke that we can hold a contingent law-like generalization in "scientific good repute" without knowing which counterfactuals are true in virtue of it. We have complained above that Fodor fails to elaborate on the notion of "scientific good repute", and wondered why some spurious law-like generalizations do not enjoy such repute. Fodor 1991—although not explicitly referring back to his criticism of Kripke—effectively elaborates this position by discriminating more finely between those law-like, but non-counterfactual supporting generalizations that we ought to tolerate, and those we ought not. The claim, at least in so far as psychological CP-generalizations are concerned, is that we may accept a law or general proposition if its antecedent conditions figure in *other*, presumably independently established, laws where they can, at least in principle, be completed. In other words, we may accept those CP-laws that share their part of their antecedent with other laws that have good scientific standing. Thus, it may be the case that the completion of a given law CP (A & … → B) is impossible, because there simply is no C_B that could do the job. Nevertheless, the law is legitimate because it is member of a set of laws

$$CP\ (A\ \&\ C_D \rightarrow D),\ CP\ (A\ \&\ C_E \rightarrow E),\ CP\ (A\ \&\ C_{F\&G} \rightarrow F\ \&\ G)\ \ldots$$

where A is successfully paired up with other completers so as to nomologically necessitate the consequent. Fodor sums up his position as follows: '… what distinguishes cp laws from other kinds of laws is that cp laws can have (real, nonrandom) *mere* exceptions and (real, nonrandom) *absolute* exceptions. What distinguishes cp laws from *anything goes* is that they can't have ACROSS THE BOARD absolute exceptions. You'd get an across the board exception iff A had a realizer which was an exception to most-or-all the laws in its network' (Fodor 1991, p. 27).

Does this take care of the case of the psychological laws that govern our arithmetical activities and their concomitant mental states? Fodor says that such laws are CP-laws which 'idealize to unbounded memory,' i.e. unbounded memory is a necessary part of conditions C_D which, together with his state of meaning *plus* by '+', enable a cognitive agent such as Don to satisfy the event type 'If queried "$m + n = ?$," responds with the sum of *m* and *n*.' On Fodor's account, this putative CP-law should encounter either just mere exceptions, or absolute exceptions that are not "across the board"—yet, this is far from clear.

2.5.2 Absolute Exceptions, Impossible Completers, and Scientific Reputation

We shall focus on one general and pervasive difficulty with Fodor's view, as well as on a simple and straightforward counterexample, the latter showing all signs of being conclusive. We turn first to the difficulty, for it will point us towards a way of salvaging if not the letter, then at least the spirit of Fodor's proposal from the counterexample. It concerns Fodor's notion of a completer. Schiffer pointed out that the mental state referred to in a psychological law's antecedent and its corresponding completer need to 'cohere', in the sense of being able, jointly, to be causally sufficient for the consequent (Schiffer 1991, p. 5). Not all things cohere, and not all properties mix to be jointly causally sufficient for something. Two electrically charged bodies can approach each other so as to be jointly causally sufficient for the occurrence of a spark, and they can associate with many other entities to cause many other things—but an electrical charge and, say, phlogiston, or the set of all objects having once been misplaced by Julius Caesar, cannot be jointly nomologically sufficient for anything. They do not mix very well. The former's existence is not accounted for by the laws of physics, and the latter's abstract nature seems to preclude it from having a causal role in the first place. Schiffer harbours principled doubts whether *all* realizers of a mental state could cohere with the completer of a given psychological law-candidate. For, on the functionalist story, a mental state can in principle be realized by indefinitely many different things—even, famously, by Swiss cheese—and if just one or a few of the mental state's realizers cohere with completers, then we do not have law-like generality (Schiffer 1991). This seems to imply that we better not take into account *all* possible realizations of a mental state. According to Schiffer, '… to say that it's a law that As cause Bs when C is to say that there is no nomologically possible world in which A-&-C obtains but B doesn't. Consequently, we need only to look at realizations of A whose conjunction with C is *nomologically possible'* (Schiffer 1991, p. 6n; schematic letters adapted, emphasis mine). A good part at least of our psychological CP-laws, like the law governing my arithmetical behaviour, do have antecedents whose completion requires the nomologically *im*possible, however. This plays of course into Schiffer's hands, who wants to deny precisely that there are any psychological CP-laws.

The claim that we need to look only at the nomologically possible when attempting to determine whether it is a law that A causes B if C, is clearly too strong, however, for it rules out from lawhood well-established scientific statements. Take the Ideal Gas Law again, which is defined over impossible conditions. Schiffer's position on lawhood suggests that what the law says is that there is no nomologically possible world in which certain conditions C obtain, and where an increase in temperature is *not* followed by an increase in pressure. *Prima facie* at least, this cannot literally be what the Ideal Gas Law says, for, surely, it is at least not *trivially* false. However, it follows trivially from the nomological impossibility of conditions C included in the antecedent (and not, as we would prefer it, from the particular way in which the antecedent and consequent are causally or otherwise related), that there cannot be a nomologically possible world where both the antecedent and consequent obtain.[97] Fodor thus has a point in insisting that this circumstance does not immediately obviate the gas-laws.

In fact, one should think that if Fodor's solution is to work at all, the notion of a completer *must* be allowed to specifically include nomologically impossible event-types, or conditions. (Fodor speaks of event-types, but I shall keep the discussion neutral between an event-ontology and a property- or state-ontology, and speak indiscriminately of 'events' or 'conditions'). Suppose completers were restricted to what is physically possible. Suppose further that the relevant mental state A in the antecedent of our putative law is 'means *plus* by '+.' By Fodor's own admission, for this antecedent to issue in the right event type B, agent D needs to dispose of unbounded memory, a condition which would be part of the completing condition of this antecedent. Without doubt, unbounded memory is, in humans at least, physically impossible. Thus it is likely that in the actual world with finite cognitive agents, this CP-law will

[97] The fact that, on the face of it, our laws are not about the actual world, but an idealized one, has lead authors such as Cartwright to argue that "the laws of physics lie" (Cartwright *How the Laws of Physics Lie*.)

face many absolute exceptions. Therefore the law fails to satisfy Fodor's condition (*i*). It seems that it would also fail condition (*ii*), however, for if our notion of a completer were restricted to what is physically possible, then the other CP-laws in the network for 'means *plus* by '+' could arguably not be completed either. (After all, in order to distinguish between behaviour issuing from a 'means *plus* by '+''-state or event, from behaviour issuing from 'means *quus* by '+'', a law would certainly need to contain physically impossible conditions in its antecedent.) Most or all laws in the network would therefore face absolute exceptions, and we would have "across-the-board absolute exception."[98]

The fact that the completer will often need to be a nomologically impossible event-type creates complications, which are tied to another kind of worry that Fodor is exercised by. Fodor worries about circularity: 'The objection would be that the story I've been telling about absolute exceptions to cp laws presupposes the unanalyzed notion of a network of laws, and thereby borrows with one hand what it spends with the other' (Fodor 1991, p. 31). Networks of laws can contain both CP-laws and strict laws, because there is no restriction on the vocabulary in which the laws in the network are couched. The same network could contain a series of psychological CP-laws as well as, for example, strict neuro-physiological laws (higher-level laws and their lower-level implementations). In fact, it could contain laws from any science whatsoever, as long as they are "about" A in the sense of referring to A in their antecedents.

Fodor toys with the idea of simply stipulating that CP-laws express one kind of nomological necessity, and strict laws another, but eventually endorses a potentially less tendentious way of pulling the rabbit out of the hat. He uses a recursive definition as follows:

> A cp law can have absolute exceptions if *(i)* it belongs to a network of strict laws for most or all of which its realizers have completers; or *(ii)* it belongs to a network of strict laws and laws that satisfy *(i)* for most or all of which its realizers have completers; or *(iii)* it belongs to a network of strict laws and laws that satisfy *(ii)* for most or all of which its realizers have completers; (etc.) (Fodor 1991, p. 32-33)

Notice that in order for the definition to work, the presence of strict laws in the network is necessary, for otherwise condition (*i*) and all subsequent ones would fail to be

[98] Nevertheless Fodor sometimes writes as if he did assume that completers are nomologically possible. See e.g. Fodor "You Can Fool Some of the People All of the Time", p. 32, and *passim*.

fulfilled. This immediately raises the question what would happen if it turns out that there are no strict laws at all, for which case Fodor provides a *caveat* with which we will deal below. Let us grant first that there are strict laws.

Given that for many networks, the completers in the network must be nomologically impossible, the strict laws in these networks will have nomologically impossible conditions C in their antecedent. Fodor's account thus seems to involve him in the interesting assumption that laws in the network can have nomologically impossible conditions in their antecedent and nevertheless remain strict, i.e. "nomologically guarantee" their consequent. Fodor might find this inconspicuous, for, he might ask, isn't this exactly what happens in the case of Boyle's ideal gas laws, where the term 'ideal gas' is used in the antecedent of the law although it has no referent in the actual world, and *idem* for countless other strict laws that employ idealizations? Well, this depends on what it means, exactly, for a law to be strict. Take the law of Supply and Demand. As demand increases and supply remains constant, the price will rise, but not inexorably so, unless competition is perfect and there is no interference—a nomological impossibility for actual markets. It seems that we must conclude that the law of Supply and Demand strictly speaking applies only to ideal markets, but not to real ones, and similarly that the Gas Law strictly speaking governs *not* gases in the actual universe, but only gases in a possible one. And, hence, that in the actual universe all these laws, whether soft or hard, apply only *non*-strictly. This would mean that strict laws defined over ideal conditions do not, after all, nomologically guarantee their consequents *in this world*, and that they are not strict after all (they are strict elsewhere).

To be strict, intuitively, is to allow no exceptions. Fodor's way of putting it is that strict laws are such that an instantiation of their antecedent nomologically necessitates their consequent (Schiffer 1991, p. 1, says that what a strict law claims is that 'A events *always* cause B-events'). On this sort of view, a law of the form $\forall(x)[F(x)\rightarrow G(x)]$ is strict if and only if there are no instances of $F(x)\&\neg G(x)$ (Fodor 1991, p. 21). But, of course, if the antecedent $F(x)$ is found to consist of events $E(x)\&C(x)$, where $C(x)$ is nomologically impossible in the actual world, then $F(x)$ will never be instantiated in that world. Obviously, we do not want to say that every law-like statement with impossible antecedents is *vacuously true* in the actual world. This would play into the hands of those who propose spurious CP-laws and ascribe spurious dispositions. A strict law that is true of *this* world says "if A, then it is not nomologically possible that ¬B". Things get interesting when A is not nomologically possible in this world. In such a case, the law looks rather like a statement about what strictly follows upon A in *another* possible world, not this one—but then, to use Fodor's words, God only knows what is consequent upon what in a nomologically impossible world. Moreover, idealized laws such as the ideal gas law are not strictly true even in a world that satisfies the antecedent conditions. As Fodor himself acknowledges, even in a world where there are ideal gases, it is true only *ceteris paribus* that an increase in pressure causes an increase in temperature (Fodor 1990, p. 95).[99]

Laws that are both strict and idealized are strange creatures, for they seem to be are literally true of, and apply strictly to nothing but, a *model*, i.e. a highly abstract and usually very simplified representation. The model necessarily leaves out of consideration countless real-world features of its object—if it did not do so, laws true in the model would not apply strictly. A model is not to be assimilated to a possible world, because unlike the former, we cannot fully specify all relevant features of the latter. In attempting to describe a given possible world, we always "leave out too much"—possible worlds are *total* ways for things to be, and any description of them that we could give would fail to be total. Hence we cannot guarantee (except by stipulation) that the Ideal Gas Law would apply strictly in any given possible world. Now, Schiffer's take on laws is that we should interpret them as talking about nomologically possible worlds, not models, a proposal that apparently avoids the difficult question of how laws that are true of an idealized model are to be brought to bear on the real-world object of the model. But does this mean that he seriously suggests that we accept only laws containing no nomological impossibilities in their antecedent, i.e. that we accept only non-idealized laws? This brings us back to the problem mentioned at the outset of this Section: given a CP-generalization of the form $\forall(x)(F(x) \rightarrow CP\ G(x))$, and assuming that F describes a physically impossible state or event type and is therefore never instantiated in the actual world, what, if anything, should we consider a confirming instance in the actual world of the generalization? If we chose to view the generalization as one that describes the actual world, albeit in an idealized way, how can it be tested? Clearly, idealizations cause serious problems for traditional accounts of confirmation, as well as for covering-law theories of explanation, and hence for traditional views (Fodor's?) of how "scientific reputation" is established.

---

[99] This has implications for our ability to describe the conditions, if any, under which ideal laws are literally true. It would seem that ideal conditions are not, after all, always completely and precisely stateable, if what we mean by "ideal conditions" are those conditions under which idealized laws are strictly true.

Fodor 1991, to resume, avails himself of a dichotomy between strict and non-strict laws without making that latter distinction clear. In particular, he implicitly presumes that, but does not explain how, strict laws can contain ideal initial conditions and nevertheless apply strictly. Fodor does not say whether he accepts the apparent consequence that laws that contain ideal initial conditions can only be strictly true of an idealized model, not the real world. How does all this tie up with our worry about completers? Very immediately: the availability of completers is what turns any given CP-law into a strict one, and our difficulty with specifying what the strict–non-strict dichotomy amounts to must be related to what, if anything, we will consider a legitimate completer. Fodor's own remarks on what it is to be "strict" are insufficient for establishing whether any law with nomologically impossible antecedents is strict, for in order to verify whether the impossible antecedent condition indeed nomologically guarantees its consequent *outside* the model, we first need to know what, if anything, we may consider its instantiation outside the model. In other words, to know specifically whether laws defined over ideal conditions are strict or CP, we need to know more about how such laws relate to real conditions, how they can be tested, and how they explain.

A full account of CP-clauses along the lines of Fodor's cannot dodge this question. For, suppose we took the stance that no actual event-type could count as a possible exception to a strict law, because *all* strict laws are idealized and do not, properly speaking, apply to the actual world at all. Then there will be, *prima facie* at least, no criterion for distinguishing strictness from non-strictness in the actual world, according to our understanding of being strict as being exceptionless. All laws true of this world will necessarily be CP. Now, if a law applies only in the conditions stipulated in the theory (i.e. in the model given by the theory), then we may of course agree that all laws defined over ideal conditions are strict. This, I take it, is the gist of Fodor's remark that the only counterfactual we can, and need to, have any confidence in is the one the theory *itself* (i.e. the model) tells us is true. Under conditions C *given in the model*, A → B. But this does not yet successfully ward off charges of circularity: if you want to explain what makes CP-laws as opposed to srict laws non-vacuously true *in the actual world*, not in the model, you need to first make clear what it is for a law to be strictly true in the actual world, and explain why apparently endemic exceptions to strict laws ought not to be counted as such.

Fodor, I hasten to say, is quite aware of the problem, but does not want to get entangled in the issue of strictness vs. non-strictness. He adds a caveat designed to deal with the eventuality that the distinction turns out to be a false one, because there are no strict laws (true of the actual world): '"Aha, but what if Hempel's[100] right and there aren't any strict laws. Then clause (i) is never satisfied, so how does the iteration ever get off the ground?"[101] His solution is: "If there are no strict laws, amend clause (i) so that it's satisfied by networks of laws which have mere exceptions"' (ibid.). This *may* have been sufficient were it not for Fodor's unsatisfactory treatment of the issue of completers. For, Fodor is oblivious to the fact that some completers need to be nomologically impossible, and that we need to distinguish those that are acceptable from those that aren't. Failure to do so obviates the 'mere' vs. 'absolute' distinction, which, in turn, invalidates his caveat and threatens CP-laws with vacuity all over again.

To illustrate this, let us consider a version of Zeno's tale of Achilles racing the turtle. Unlike in Zeno's paradox, the turtle does not get a head start. Suppose you are observing them as they approach the start line, and you make the following sensible prediction: "All other things being equal, Achilles will outrun the turtle." I disagree and say: "All other things being equal, the turtle will outrun Achilles." Both statements may be seen as vouched for by corresponding specialised biological CP-laws, namely,

> (1) CP ∀(x)(if *x* is an adult member of *Homo Sapiens*, *x*'s top running speed exceeds that of adult members of *Reptilia Testudines*)
> (2) (1) CP ∀(x)(if *x* is an adult member of *Reptilia Testudines*, *x*'s top running speed exceeds that of adult members of the species *Homo Sapiens*)

On Fodor's story, if Achilles does not in actual fact outrun the turtle, then (1) will have encountered a mere exception. (2), on the other hand, is obviously *spurious*. It is one of the CP-law candidates we must exclude in order to save CP-clauses from vacuity. (2) should thus either fail Fodor's condition (*i*), because there is no enrichment of the antecedent such that the antecedent together with the laws of nature necessi-

---

[100] Fodor refers to Hempel, C. G. (1988). "Provisoes: A Problem Concerning the Inferential Function of Scientific Theories" *Erkenntnis* **28**: 147-164. According to Earman and Roberts ""Ceteris Paribus", There Is No Problem of Provisos", Fodor very likely misunderstood Hempel here, who did not wish to deny that there are any strict laws. The authors point out that Hempel uses an idiosyncratic sense of "proviso" that is weaker than that of Fodor's 'CP'-clause. The right interpretation of Hempel, according to Earman and Roberts, is that he acknowledges that even strict laws need provisos (in Hempel's sense), but denies that they are *ceteris paribus* (cf. Earman and Roberts ""Ceteris Paribus", There Is No Problem of Provisos", p. 445). Provisos, in Hempel's usage, are conditions on the validity of the *application* of a theory containing strict laws to a given physical system, not conditions on the truth of these laws themselves (cf. *infra*).

[101] Fodor "You Can Fool Some of the People All of the Time", p. 33.

tates the consequent, or condition (*ii*), because the network of laws it is a member of would fail to contain such a C as well, making the exception across-the-board.

It seems, however, that (2) does not fail Fodor's test. Suppose that what actually happens is that Achilles and the turtle start the race, and, miraculously, the turtle wins. Achilles is unwell and can barely move. Fodor would say that Achilles did not succeed in overtaking the turtle because the necessary completer C_A of CP-law (1) was not instantiated. Plausibly, part of the completer of (1) as applied to our example is the condition that Achilles is in a state of relative health. In the actual event, however, initial conditions were that Achilles was seriously indisposed from the previous nights' celebrations. C_A, a very ordinary and nomologically possible condition, just happened not to be the case, and we have a mere exception that does not render the CP-statement false. So far so good. However, suppose we repeat the race a few days later, and now Achilles wins. Unwilling to acknowledge that my CP-law candidate has encountered a counterinstance, I argue, in an exactly parallel manner, that the turtle didn't run faster because a part of the relevant completer C_T was not instantiated. Suppose that part of C_T is a species of "cosmic weather," bursts of an unknown kind of cosmic radiation emanating from a massive black hole at the centre of the galaxy. The effect of C_T-radiation on turtles are multiple, ranging from muscular stimulation and oxygen enrichment of the blood, to greater flexibility of joints, etc. In short, the effect of C_T-radiation on turtles is that when subjected to it they outrun any terrestrial being.[102] However, the physiological effects of C_T-radiation persist only as long as the radiation itself, which although usually a rather permanent feature of the weather in the Milky Way, is subject to random variation. As it happens, during the cosmologically brief time span in which humans have engaged in the scientific observation of nature, there was a calm with respect to C_T-bursts, which is why no one has yet seen a turtle overtake a man. This seems to take nothing away from the fact that, in the greater scheme of things, CP(turtles overtake humans). What has gone wrong? Has anything gone wrong?

John Earman, John Roberts, and Sheldon Smith[103] describe the difficulty as follows:

[102] The scenario is inspired by Boghossian, P. A. (1989). "The Rule-Following Considerations"Ibid. **98**: 507-549, at p. 529-30, who writes '... not every true counterfactual of the form 'If conditions were ideal, then, if C, S would do A' can be used to attribute to S the disposition to do A in C. For example, one can hardly credit a tortoise with the ability to overtake a hare, by pointing out that if conditions were ideal for the tortoise—if, for example, it were much bigger and faster—then it would overtake it.'
[103] Earman, J., J. Roberts, et al. (2002). "*Ceteris Paribus* Lost" *Erkenntnis* **57**: 281-301.

> It seems that there could be no informative account of the truth-condtions of a CP law-statement that did not render them vacuous. One way to see the problem is to note that we could specify the conditions under which such a statement is true if and only if we could specify the conditions under which it is false, but that is exactly what we cannot do with a CP law-statement. For such a statement will be violated exactly when the regularity contained in it is violated and "other things are equal", i.e. there is no "interference". But we cannot specify the conditions under which the second conjunct obtains; otherwise the CP clause is simply an eliminable abbreviation and what we have is not a genuine CP law-statement. (Earman, Roberts *et al.* 2002, p. 292).

Now, the purported interferer to my CP-law, the physiological effects of C_T-radiation on turtles, is extremely *ad hoc*, and likely to be physically impossible. Even so, Fodor allows completers to be nomologically impossible, and C_T is certainly not *logically* impossible. In terms of the traditional deductive-nomological account of explanation (e.g. Hempel 1966, pp. 47-69), in the Turtle example I have simply accounted for what otherwise would have been a clear disconfirming instance of the law-candidate by holding that during the race the requisite background conditions were not satisfied. A description of these rather exotic conditions is a part of my set of *explanans* sentences that, together with my CP-law, deductively imply the *explanandum* sentence describing the event of the turtle overtaking Achilles. The point of the example then becomes that posing no restrictions whatsoever on completers is tantamount to posing no restrictions on plausible background conditions. As Earman, Roberts *et al.* 2002 point out, the CP-clause cannot perform this job for us, as it does not (by its very nature) exhaustively describe the conditions it refers to, and hence it excludes next to nothing. On the other hand, our background knowledge, or auxiliary hypotheses, cannot perform the job either. For, if one of these hypotheses included the claim or its equivalent that "there are no interferences", then this hypothesis too would have to be stateable in a form such that it allows us to check whether it is true or not (Earman, Roberts *et al.* 2002, p, 293). They note that if this is not the case, then '... the prediction relies on an auxiliary hypothesis that cannot be tested in itself. But it is generally, and rightly, presumed, that auxiliary hypotheses must be testable in principle if they are to be used in an honest test. Hence, we can't rely on a putative CP law to make any predictions about what will be observed, or about the probability that something will be observed. If we can't do that, then it seems that we can't subject the putative CP law to any kind of empirical test'. (Earman, Roberts *et al.* 2002, p. 293).

So, on Fodor's account "CP, the turtle will outrun Achilles" did *not* encounter an absolute exception because the antecedent does have a completer. True, the particular completer dreamt up here looks extremely implausible, and will probably be physically impossible not only in *this* world, but also in many other worlds remotely similar to ours. But Fodor's specification that completers are 'arbitrary event types' is not a constraint, but a wild card. Taken literally, it does not even rule out logically impossible events, unless one considers logical possibility to be built into the concept of an event. If we can have access to an unrestricted class of nomologically impossible events or conditions to complete the antecedent states of our putative CP-laws, then, surely, we will be able to find a completer for *any* antecedent state. As David Lewis sometimes puts it, *anything can cause anything.* This means that *any* putative CP-law would fail to encounter absolute exceptions and all counter-instances, even if rather endemic in the actual world, could be explained away as 'mere exceptions.' If they can be so explained, however, then CP-laws would not be empirically testable. According to Earman *et al*., this fact, beyond any semantical worries concerning a possible indeterminacy in the *meaning* of a CP-clause, really is the Achilles heel of any philosophy of science countenancing CP-laws.

> In order for a hypothesis to be testable, it must lead us to some prediction. The prediction may be statistical in character, and in general it will depend on a set of auxiliary hypotheses. Even when these important qualifictions have been added, CP laws still fail to make any testable predictions. Consider the putative law that CP, all Fs are Gs. The information that x is an F, together with any auxiliary hypotheses you like, fails to entail that x is a G, or even to entail that with probability p, x is a G. For, even given this information, other things could fail to be equal, and we are not even given a way of estimating the probability that they so fail. (Earman, Roberts *et al.* 2002, p. 293).

The other things that failed to be equal, in our Turtle case, concerned the absence/presence of C_T-radiation, and clearly, we have no way of estimating the probability of C_T-radiation and its effects on turtles being real.

Earman, Roberts, *et al*. have already pre-empted the most obvious objection at this point, the appeal to background knowledge: the purported effects of C_T-radiation on turtles, so the objection goes, are flatly ruled out by everything else we currently know about the physics of our universe and the biology of creatures on Earth, and there is scarcely more point to speculating about what would happen if there was any such radiation, than there is to calculate the consequences of Jupiter being made of

Swiss Cheese, given that everything we know suggests that it is not made of it. In Fodor's terminology, the completer is unacceptable. Even if Earman, Roberts, *et al*. are correct and CP-laws are instrinsically untestable and hence ought not be admitted, we might nevertheless want an account on which certain CP-statements are strictly frivolous, whereas others are of empirical interest, e.g. because they might lead research towards truly strict laws. The problem is that an argument strictly parallel to the way in which we have rejected the turtle-law above says that perfectly elastic collisions and impermeable containers are *also* ruled out by the physics of our universe, and that there is little point in speculating about what would happen *in this world* if there were any such things as these. All our background knowledge suggests that nothing in this world ever collides 100% elastically. If there were any such things in this world as perfectly elastic collisions between molecules, then, plausibly, some objects would never cool down, other objects would never stop moving—we could build a *perpetuum mobile!*—and countless other macroscopic objects would behave in a way radically different than what we are used to. It would be a very, very strange place, little less stranger than a place in which turtles overtake humans. Indeed, it would arguably be a lot stranger than that.

How can we distinguish between the Turtle case, and OK idealizations about ideal gases? There is, to be sure, a *de facto* scientific difference between CP gas laws and our purported CP law about turtles' running capacities to which I shall come in a moment—but is there any *philosophical* difference? Fodor provides no indication here. He does not explain why we should prefer the idealizations contained in the former (impermeable containers), over the ideal conditions of the latter (C_T-radiation). One idea would be to rely on the premise that certain nomologically impossible things are *less possible* than others, C_T-radiation and its effects on turtles being less possible than impermeable containers, and that when using idealizations to make predictions about the behaviour of things in this world, we should use those that are nomologically less impossible. This looks rather hopeless. I take it that while there may be a (mildly convincing) argument for the view that given that it is nomologically impossible to travel faster than light, it is *less* impossible to travel 1% faster than light as it is to travel 200% faster, there is no convincing method whatsoever for ordering *qualitatively* different impossibilia. Is it more or less impossible for humans to live one billion years, than it is for one of them to jump from here to the moon? Even serious proponents of a relation of overall similarity between possible worlds, and of the existence of a partial ordering of worlds with respect to their proximity to the actual world, acknowledge that the concept must remain vague—David

Lewis first and foremost. We may safely presume that an attempt at ordering impossibilia would not help us decide questions such as ours. Idealization is not a matter of (quantitative) approximation of the actual, and sometimes it seems that it is not even a matter of approximation at all (see *infra*).

Before I come to the most obvious way of approaching the problem, the argument from scientific predictive/explanatory success, a word about a more desperate retrenched position. One might argue that I have illicitly expanded Fodor's account to a domain it was not designed for. After all, Fodor meant to cover psychological CP-laws about multiply realizable propositional attitudes only, and the Achilles example is a physical one. Our point about impossible completers generalizes, however. Unbounded memory is also a biologically impossible condition for human agents, and it is, arguably, equally as impossible as other nomologically impossible conditions, be it C_T-radiation or something else. The point is: if unbounded memory is permitted, then why not unbounded force, vision, hearing, speed, etc.? According to Fodor,

> CP (Don means *plus* by '+' → Don is able to reply to every query of the form 'm+n = ?' with the sum of m and n),

is legitimate, because we can enrich the antecedent as follows:

> CP (Don means *plus* by '+' & Don has unbounded memory → Don replies to every query of the form 'm+n = ?' with the sum of m and n).

Fodor would of course acknowledge that the antecedent needs still further enrichment before we could actually strip away the CP-clause. But even if completion was impossible, the law could still qualify on Fodor's story if there is a relevant means-*plus*-by-'+'-network of completable laws. Unfortunately, this does not explain why I am I *not* also allowed to assert

> CP (Don intends to swim → Don swims faster than a dolphin)

After all, part of the completing conditions for that antecedent could be:

> CP (Don intends swim & Don has unbounded muscular strength → Don swims faster than a dolphin)?

Again, the antecedent needs further enrichment—e.g. presence of a suitable body of water, Don is not sick, etc.—and may not be completable, but we have seen no reason to reject the idea that if there is a qualifying means-*plus*-by-'+'-network of laws, then there could also be a qualifying intends-to-swim-network.

The best (only?) option at this point clearly is to take guidance from existing scientific practice, in other words, Fodor's appeal to 'scientific reputation'. The CP-law about meaning is admissible because there is, for the relevant antecedent, a network of established true laws. Fodor is not explicit, in the case of 'means *plus* by '+'', about which laws he has in mind. Let's assume it is linguistic laws, i.e. laws from a science which begins to enjoy explanatory as well as predictive success. The problem with this reply, within the framework provided by Fodor, can be stated immediately, and it simply reiterates our above difficulties with completers. On Fodor's theory, the existence of a network of valid laws with the same antecedent—or what I take in this context to amount to essentially the same thing, "scientific reputation"—is *unnecessary* if condition (*i*), i.e. availability of a completer, is already satisfied. But some completer or other, no matter how bizarre, will always be available, as we have seen. Moreover, if for any CP-law we can find a completer to make it comply with condition (*i*), then we can, of course, make the law comply with condition (*ii*). In fact, with an overabundance of possible completers at our disposal, we can make any body of pseudo-laws, say the "laws" of astrology or parapsychology, comply with Fodor's conditions for being acceptable CP-laws.

This is not to suggest that our CP-law candidate about turtles is unjustly ignored by biologists, or that the hypothesis that humans can indeed swim faster than dolphins has somehow escaped legitimate attention. Science has its own inexplicit rules for hypothesis selection, and philosophers of science are perhaps well advised to leave them as they are. Individual sciences do not have a general theory of CP-laws—probably because they do not need one—and we are not currently in the business of providing one for actual use in hypothesis selection. However, it is a serious weakness of Fodor's explicitly *philosophical* theory, a theory designed to show why CP-laws are not vacuously true, that it lets our spurious candidates slip through the net. A running repair is needed to impose constraints on the type of completing event-type allowed. Taking account of our discussion so far, we might try a proposal such as CP(A→B) is true if and only if either of the following conditions is satisfied

a) $A \rightarrow B$ is strictly true
b) if R_i realizations of A are exceptions to $A \rightarrow B$, then there must be a completer C_B for $R_i(A)$ such that there are *other* strict laws of the form $E \,\&\, C_B \rightarrow D$, $F \,\&\, C_B \rightarrow G$, $H \,\&\, C_B \rightarrow K$, etc.

Here, we have simply dropped the distinction between mere/absolute exceptions, as there will be no absolute exceptions to any law if completers can be anything. All laws are either exceptionless, or they are CP and have exceptions. The first condition says, trivially, that if a law A→B is strict, then adding a CP-clause to it will not render it false. Strict laws are special cases of CP-laws for which everything is always equal. The constraint on legitimate completers is implemented in (b), which says that if some realisations R_i of antecedent state A do not force B on its own, then the completed law $A(R_i)$ & $C_B \rightarrow B$ will only count as a law (or as a serious law-candidate requiring empirical confirmation) if the relevant completer C_B also occurs in the antecedents of other laws, which may be part of other sciences, with different consequents D, G, K. Thus, the assumption of unbounded memory is sometimes made in generative linguistics for central theoretical claims about an ideal native speaker's linguistic competence.[104] Unbounded force, on the other hand, is not an assumption routinely made in *any* science, nor is of course the existence of C_T-radiation. It helps to predict nothing, and explains even less. So, if unbounded memory is an accepted idealization in linguistics, we seem justified to at least provisionally accept unbounded memory as an idealization in psychology as well. (We may do so, perhaps, until the usefulness, etc., of the idealization is somehow established within psychology itself, e.g. through the predictive and explanatory successes of theories built on it). *Prima facie* at least, our amended criterion would seem to exclude any "laws" that appeal to unbounded force, etc.—for there is no network of laws which make the assumption of unbounded force, or of C_T-radiation, for that matter, in their antecedent.

The amendment is still unsatisfactory, however. Most obviously, the proposal, just as Fodor's initial stance, presupposes the availability of strict, exceptionless, laws governing the behaviour of a given antecedent event-type or condition, and thus risks excluding CP-laws about antecedents for which such laws are unavailable. For example, although there may be an existing network of laws in linguistics that appeal to unbounded memory, these laws actually have little chance of being strict. Nor is there

[104] It is quite standard to encounter statements in linguistics texts such as: 'We can, in principle, determine the grammaticality of sentences that are arbitrarily long and complex (though, this might be impossible in practice, because of fatigue, memory limitations, and limited life times'; or: 'Theoretical linguistics is concerned primarily with developing an adequate theory of linguistic competence in ideal situations, that is excempt of external factors which may affect production or perception such as false starts, hesitations, memory lapses, repetitions, etc.' (Prinz, J. J. (2001). "Philosophy of Cognitive Science: Chomsky's Linguistics", http://www.artsci.wustl.edu/~jprinz/cogsci2001/cog2001-3.htm. (accessed 25/4/03), and (Amores, J. G. (2002). "*Morfosintaxis Inglesa.* Unit 2: From Taxonomic to Generative Grammar", http://fing.cica.es/~gaby/Docencia/Morfo301/Morfo301.htm. (accessed 25/04/2003), respectively.)

likely to be any other science, which could provide us with a network of such strict laws. Moreover, on the amended story we cannot provide for the eventuality that there are no strict laws by following Fodor and adopting something like his recursive definition above, because that definition is predicated on the mere-absolute exception distinction. Failing availability of such a distinction some expression such as 'law with exceptions,' or 'CP-law', would occur both in his *definiendum* on the left as well as in every recursive condition in the *definiens* on the right.

However, the improved analysis does have the virtue of escaping a simple but decisive counterexample in the literature, which shows that there are indefinitely many *prima facie* spurious generalizations that satisfy Fodor's conditions for being true CP-laws. The counterexample that shall be presented here is an adaptation of objections by Pietroski and Rey and, independently, Gerhard Schurz,[105] to a recent alternative theory of CP-laws.

2.5.3 Completers and Independent Explainers

The alternative theory is given by Pietroski and Rey 1995, who develop an account of CP-laws consonant in many ways with Fodor's. Fodor's underlying idea was a simple and intuitive one. The reasonable thing to assume when confronted with Kripke's Sceptic is that most normal speakers mean *plus* by '+', whereas the *quus*-hypothesis is evidently spurious. Surely, if the assumption that all or most of us mean *plus* by '+' is true, then this mental state ought to be subsumed by a great number of true psychological laws. For, according to Fodor '… if a kind of state figures in any laws, then it will figure in many.'[106] On the other hand, there should not be any valid empirical generalizations about quirks of nature. A philosopher who believes in objective regularities, and laws of nature that are descriptions of these regularities, must assume that although things may sometimes go wildly wrong, as in the hypothetical *quus* case, they cannot *regularly* go wildly wrong—and when things do go wrong, they do not always go wrong in the same way. Hence Fodor's call for a network of valid law-like generalizations about normal agents who mean *plus* by '+', for there ought to be nothing comparable for those agents, if any, who mean *quus*. Hence also

---

[105] Pietroski and Rey "When Other Things Aren't Equal"; Schurz, G. (2001). "Pietroski and Rey on Ceteris Paribus Laws" *British Journal for the Philosophy of Science* **52**(2): 359-370.
[106] Fodor "You Can Fool Some of the People All of the Time", p. 27.

his suggestion that if things do always go wrong, i.e. if a putative law CP(A→B) encounters only absolute exceptions, then there is no real instance of A→B to generalise upon in the first place, and A→B does not represent a genuine regularity. It is simply false (Fodor 1991, p. 26).

Pietroski and Rey depart from a very similar intuition, for they too believe in lawful regularities. What we literally mean when we say 'All other things being equal, X will do Y', is simply that X will do Y if nothing "goes wrong," where things go wrong when something *interferes* with X's doing Y. 'CP (A→B)' is thus to be construed as "A→B, unless something interferes," and Pietroski and Rey purport to show under which conditions such sentences are non-vacuous (not under which they are true). But whereas Fodor chose nomological necessity, or lawhood, as his primitive notion, they use 'explanation.' Their essential idea is this: if we dispose of an independently elucidated notion of 'X explains Y,' then a CP-law of the form

$$CP\ [\forall(x)(F(x)\rightarrow\exists yG(y))]$$

is non-vacuous if and only if for every instance that represents an apparent exception to the law, i.e. for every case of F(x) & ¬G(y), there is a fact distinct from F(x) such that that fact (a) explains ¬G(y), and (b) is 'explanatorily independent' from ¬G(y).[107] Explanatory independence, in turn, is to be thought of as a relation between X and Y such that X is independent from Y if and only if there is a fact Z explained by X, but Z is neither an analytical consequence of Y, nor causally dependent on the occurrence of Y (Pietroski and Rey 1995, Ibid.). Pietroski and Rey introduce 'explanatory independence' in order to exclude all *ad hoc* explanations of why things have gone wrong that are available to us when a putative CP-law encounters an exception.

They illustrate with the following case: suppose someone wants to defend the thesis that certain humans have extraordinary psychic powers by postulating some such CP-generalization as

> CP (if parapsychologiclal agent X predicts *p*, then *p*) (Pietroski and Rey 1995, p. 90)

[107] For the full semi-formal statement of their sufficient condition for CP-lawhood, see Pietroski and Rey "When Other Things Aren't Equal", p. 92.

Sober scientists will be quick to point out that this "law" is bound to encounter many cases of X predicts *p,* and subsequent instances of $\neg p$. These could conceivably be explained away by the proponent of the law. She might, for every occurrence of $\neg p$, appeal to a hitherto unknown interfering factor, namely "hectoplasmic interference." As we have already seen with our example of "C_T-radiation", the same sort of move is available to save *every* putative CP-law imaginable. '… if 'cp F => G' means merely that 'F => G' is true in those circumstances in which there are no instances of F and not G, then 'ceteris paribus-laws' look to be strictly tautologous — true, but presumably not explanatory laws in an empirical science' (Pietroski and Rey 1995, p. 87).

Pietroski and Rey's requirement that interfering factors be explanatorily independent is the demand that whatever explains the apparent exception also have an explanatory life of its own, i.e. that it does some explanatory work that "has nothing to do" with the relevant exception. CP-laws are 'cheques written on the banks of independent theories,' they say, in the sense that every accepted CP-law entails a commitment that there must be successful independent theories which subsume the apparent exceptions and use them to their own explanatory ends. If hectoplasmic interference is to be real, it must succeed in explaining other things as well, not only the failure of parapsychological experiments. *Ditto* for C_T-radiation.

It is plausible that some such principle is at least one of the implicit rules governing scientific hypothesis selection, and it seems that Pietroski and Rey have succeeded, at least to a first approximation, in capturing this principle in a philosophically perspicuous form. We can also see the similarity between Fodor's and Pietroski and Rey's approach: the former thought that if there is a genuine mental state 'A' such that CP (A→B), then there surely must be other true laws in which 'A' figures. This was Fodor's way of ruling out spurious states and CP-laws that might be justified by the equivalent, in his account, of hectoplasmic interference, namely 'absolute exception'. 'Absolute exception' fulfils the same role as 'hectoplasmic interference' because the fact that a given realizer state R of mental state A has no appropriate completer serves to *explain* every instance of R(A) & ¬B. However, as we have seen, Fodor's position was that if factors such as hectoplasmic interference truly exist and CP(A→B) has an absolute exception because of them, then for CP(A→B) to be genuine nevertheless there must be many other laws

$$CP(A\&C_D \rightarrow D),\ CP(A\&C_E \rightarrow E),\ CP(A\&C_F \rightarrow F),\ \ldots$$

This means that Fodor's account says nothing at all about the *culprit* at the root of all our difficulties, namely the completer C_B whose availability (though not instantiation) makes the law true. Our running repair in the previous Section was intended to elaborate on the role of the completer, and this is also, in a different way, what Pietroski and Rey do.

They require that if it is (the absence of C_B) that interferes with the factual truth of $A \rightarrow B$, then C_B must figure in the explanation of independent phenomena, and hence in independent theories (we required it to figure in independent laws). Assuming, with Fodor, that explanation at least sometimes works by invoking laws, we can put Pietroski and Rey's suggestion in terms of laws as follows: we can accept $CP(A \rightarrow B)$, even though $\neg CP(A \& C_B \rightarrow B)$, if

$$CP(A \& C_B \rightarrow D),\ CP(G \& C_B \rightarrow I),\ CP(K \& C_B \rightarrow L),\ \ldots$$

Alternatively, if we reject the covering law-view of explanation, the proposal is that we can accept $CP(A \rightarrow B)$ if

$$A \& C_B \twoheadrightarrow D,\ G \& C_B \twoheadrightarrow I,\ K \& C_B \twoheadrightarrow L,\ \ldots$$

(where "$X \twoheadrightarrow Y$" means X *explains* Y). Pietroski and Rey do not say as much, but their account amounts to a proposal as to how to lay additional constraints on a mode of inference Charles S. Pierce called "abduction", and what others call theoretical inference, or inference to the best explanation. Peirce described abduction as having the logical form of an inverse *modus ponens*, a kind of deductively invalid "reasoning backwards" from consequent to antecedent: 'The surprising fact, C, is observed; But if A were true, C would be a matter of course, Hence, there is reason to suspect that A is true' (Peirce 1931-35, Vol. 5, p. 189). The defender of Pietroski and Rey's example of a spurious law, 'CP (if parapsychologiclal agent X predicts *p*, then *p*)' is indeed faced with a "surprising new fact", from her point of view, namely $\neg p$, a counterinstance to her law. She uses abduction to explain it: if it were true that there was hectoplastic interference, then $\neg p$ would be a matter of course. Hence, there is reason to suspect that there was hectoplasmic interference. Our grounds for suspecting that the abduction, or inference to the best explanation, is unacceptable are precisely that this explanation is not the *best* available, on account of the nature of the postulated interferer. In Pietroski and Rey's eyes, whatever is claimed to interfere with X's being followed by Y must not only be *real*, there must be a successful explanatory practice based on assuming its existence. A perfectly reasonable way of specifying what it is for a putative interferer to be real and accessible to scientific inquiry is to require that its postulation allow us explain other, independent, phenomena in the world as well. A phenomenon that is not real cannot explain anything, and if the postulation of a phenomenon also explains something other than just the disputed phenomenon, we have grounds for assuming it to more than merely *ad hoc*, namely real. Inferences to the best explanation must choose the *best* explanation, and Pietroski and Rey's proposal amounts to a (at least partial) specification on what it is to be the best.

Prima facie, the account seems to work reasonably well for hectoplasmic interference, C_T-radiation, and pseudo sciences such as astrology, etc. It seems indeed to represent at least part of the story of how science deals with apparent exceptions to laws. For example, it fits beautifully the circumstances in which astronomers Le Verrier and Adams were led to postulate the existence of Neptune to account for apparent counter-instances to Newtonian physics presented by Uranus' orbit. The existence of Neptune serves to explain a whole raft of observations independent of any matters concerning Uranus' orbit. Unfortunately, however, all is not well with Pietroski and Rey's theory. Explanation of causally and analytically independent facts may indeed be necessary for being a genuine interferer in the relevant sense, but it turns out to be *insufficient*. Earman and Roberts give the following counterexample:

> … let "Fx" stand for "x is spherical", and let "Gy" stand for "y = x and y is electrically conductive". Now, it is highly plausible that for any body that is not electrically conductive, there is some fact about it—namely its molecular structure—that explains its non-conductivity, and that this fact also explains other facts that are logically and causally independent of its non-conductivity—e.g. some of its thermodynamic properties. … If Pietroski and Rey's proposal were correct, then it would follow that ceteris paribus, all spherical bodies conduct electricity. (Earman and Roberts 1999, p. 453)

If we can explain any case of ¬G(x) by appeal to a fact that is logically and causally independent of whether F(x), then for *any* F and *any* G, CP(F(x)→G(x)), which has the effect of making the CP-clause trivial again. Further examples can be found at will: if ¬G(x) stands for 'x is not a good football player', and F(x) stands for 'x is blond', then it suffices for there to be an explanation of why Peter is not a good football player that is logically and causally independent of whether Peter is blond, for it to be true that

> CP (if Peter is blond, then Peter is a good football player)

Clearly, for CP(F(x)→G(x)) to be a law, it cannot be enough that there is an independent explanation of every apparent exception F(x)∧¬G(x) (Earman and Roberts 1999, p. 454).[108] Earman and Roberts' diagnostic of the cause of the problem is that we have not made sure that F is *relevant* to G. Whether Peter is blond is not causally or otherwise relevant to whether he is a good striker, at least as far as our current understanding of these things goes, and there will surely be explanations of the level of his football skills that are logically and causally independent of the colour of his hair. Earman and Roberts point out that if Pietroski and Rey grant the counter-example and the diagnostic (which, it seems, they must), then they face the uphill struggle of giving a satisfactory specification of the appropriate notion of relevance without re-using the notion of a CP-law (Earman and Roberts 1999, p. 454).[109] Simply availing oneself of a primitive relevance-relation is not an option in the present context, for 'the kind of relevance in question is something we understand by way of our notion of law' (Ibid.). Like Fodor, Pietroski and Rey encounter their own vicious circularity problem.

In fact, given the proximity between Pietroski and Rey's proposal and Fodor's, it would not be surprising if Fodor's theory did not face the same sort of counter-example. As we saw, the role of the independent explainer (the independent theory) in Pietroski and Rey is played by the network of laws in Fodor. The question is therefore if there might not be a case in which we have a patently absurd candidate CP(A→B), for which there is nevertheless an appropriate network of A-laws. Sure enough, we need not search very far. Suppose our candidate CP-law says

CP (If *x* is spherical → *x* conducts electricity)

The antecedent does not, on its own, nomologically necessitate the consequent and needs further conditions to obtain, such as, for example, the fact that *x* is made of a material with the appropriate micro-properties. Not all spherical bodies fulfil that condition, and there will be, in Fodor's terminology, absolute exceptions to this

[108] Schurz, G. (2001). "Pietroski and Rey on Ceteris Paribus Laws"Ibid. **52**(2): 359-370, pp. 366-67, reaches exactly the same conclusion. He also argues that Pietroski and Rey's "deductivistic" approach presupposes determinism (Ibid.).

[109] Similarly, Martin, C. B. (1994). "Dispositions and Conditionals" *Philosophical Quarterly* **44**(174): 1-8, pp. 5-6, claims that it is impossible to non-trivially specify those conditions that are *relevant* to a given CP-conditional, and uses this to argue that conditionals prefaced with a CP-clause cannot provide a reductive analysis of the meaning of disposition ascriptions, for in order to know what to do with the CP-clause, we already need to understand the disposition ascription to be analysed.

"law." On his story, though, the decisive question is if it encounters absolute exceptions across-the-board. For this, there should be no '*x*-is-spherical'-network of other true laws such that "many or most of them" have antecedents that necessitate their consequents, or can be completed to necessitate them. Therefore, for the counterexample to work, it suffices if there are a number of true laws about spherical objects, which are either strict or encounter mere, but not absolute, exceptions.

Now, as we have seen, whether there are strict laws is a moot point, because we have yet to find an uncontroversial example of one. Surely, however, if there are any such things as CP-laws with mere exceptions at all, then there will be some with '*x* is spherical' in the antecedent. Take generalizations such as

> CP ∀(x)(if *x* is spherical & *x*'s centre of gravity lies at the point of suspension or support) → *x*'s centre of gravity is in neutral equilibrium) [in other words, *x* "rolls" if pushed]
> CP ∀(x) (if *x* is spherical & *x* has a higher temperature than its surroundings → *x* dissipates heat to its surroundings at a slower rate than non-spherical objects of the same material under identical initial conditions)
> CP ∀(x)(if *x* is spherical & *x* carries an electrical charge → the electrical field around *x* is indistinguishable from the electrical field around a point carrying the same charge)

The above generalizations are predicated on the fact that although many properties of a spherical body will follow analytically from the geometrical properties it shares with all spheres—such as having a surface area of $\pi(2r)^2$, or having the smallest volume/surface ratio of all solids—, there are *physical* consequences of having these geometrical properties that are contingent upon the properties of *other* things and the laws of physics. The chances of the above generalizations that capture some of these regularities of being valid CP-laws are as good as those of any CP-law candidate. *Ditto* for generalizations about what is consequent upon being blond. Therefore the counterexample works against Fodor as well.

If CP-laws are 'cheques written on the banks of independent theories' about the relevant interferers for Pietroski and Ray, then for Fodor they are cheques written on the banks of independent laws about the relevant antecedent states. The non-vacuity of a given CP-law depends for Fodor on there being other successful laws that subsume the troublesome, because incompletable, antecedent. Just as Pietroski and Rey struggle with the fact that there are too many independent explanations, the problem for Fodor is that *for any antecedent, there is no shortage of true same-*

antecedent-laws. Peter Mott,[110] who comes to substantially the same conclusion in assessing Fodor's theory of CP-laws, puts it this way: 'The point is that an arbitrary law like "If you are thirsty you will eat salt" passes [Fodor's] test parasitically on the back of the other true laws. What follows from this … [is that Fodor's test] simply fails to exclude any candidate law. It is a vacuous necessary condition, everything passes it. Fodor's account therefore tells us nothing at all about cp-laws' (Mott 1992, p. 340). Our diagnostic of the problem for Fodor ought to be the same as for Pietroski and Rey: it is not sufficient to require, for every candidate CP-law about A, a network of completable laws about A. We need a further constraint on the range of admissible laws, i.e. we need to identify the set of network laws A→D, A→E, A→F, etc., that are in the appropriate sense *relevant* to whether A→B.

Fortunately, our improvement of Fodor's account fares better with the counterexample. For, on our version, obviously false laws of the form CP(A → B), although in principle completable by a spurious completer C_B, will *not* have any legitimate network of laws of the form D & C_B →E, F & C_B → G, H & C_B →K, etc. to fall back on. For the spurious completer C_B will not be part of any recognized scientific activity. This will make CP(A → B) fail our condition (ii), that the particular completer in question be in use by some other science for the statement of other laws. Recall that Fodor criticised Kripke for unduly worrying about the apparently fantastical assumption of infinite memory when ascribing the disposition to add, by pointing out that entirely respectable sciences use similar idealizational assumptions, such as infinite impermeability and elasticity. Our objection was that Fodor gives no indication of why comparable idealizing assumptions, such as infinite strength, that license evidently spurious CP-claims are not *ipso facto* permissible. To save Fodor's argument from vacuously endorsing all CP-claims we need to show that some methods of completion are acceptable whereas others are not. Our view is that the gist, although not the precise form, of Fodor's suggestion was quite right, and that successful science is our only, if probably insufficient, guide. Perhaps acceptable CP-laws depend for their legitimacy on their deploying completers that are identical, or at least similar, to *existing* completers in successful predictive and explanatory laws. Although considerable care would be necessary to avoid making such a criterion too conservative and descriptively inadequate of scientific practice (in particular, by ruling out theory *change*), it might point the way towards a methodology for eliminating C_T-radiation, hectoplasmic interference and their likes without engaging in elaborate and wasteful

[110] Mott, P. (1992). "Fodor and Ceteris Paribus Laws" *Mind* **101**(402): 335-346.

empirical testing. For it might provide an explicit method for excluding spurious potential interfering "factors" on purely theoretical grounds. The philosophical (if not scientific) urgency of such a method is obvious given the practical impossibility of empirically disconfirming *all* far-fetched sceptical hypotheses such as the *quus*-hypothesis.

In this context the reasons for which Fodor 1991 and Pietroski and Rey 1995 fall short are instructive. These authors have quite sensibly looked to established scientific laws and scientific explanatory practice for guidance as to which CP-laws are acceptable, and what makes CP-laws non vacuous. That their proposals have not achieved the advertised goal is ultimately due to the simple fact that science is not concerned with typically philosophical sceptical arguments. Scientific practice does not wear on its sleeves explicit criteria that would allow us to formulate principles that discriminate between the legitimate and the spurious cases. For example, verification of the assumption 'CP(humans can fly)' is not on the agenda of human physiology and biology, and it is also not part of its tasks to illuminate us about its reasons for doing so—in contrast to the claim 'humans can fly', which it rejects, for sound empirical reasons. Yet, there is little within physiology and biology that tells us explicitly why the former claim is ignored, while the latter is not. Moreover, philosophy's labour to find criteria for eliminating such obviously absurd "hypotheses" may seem, at best, specious to the scientist (or typically "philosophical")—and at worst a gregarious waste of time. She will be right in so far as the potential vacuity of all CP-clauses is neither on the scientist's horizon of problems, nor indeed is there anything in her theoretical toolkit to solve it. *Philosophy* of science, however, traditionally has taken very seriously the task of demarcating what is genuine science from what is not, and there are rather good arguments for believing that this is indeed an exceedingly important *philosophical* task.[111] I see the attempt to identify the hallmarks of spurious CP-claims, with a view to distinguishing them from those that have at least a chance of being true and hence merit empirical investigation, as part of that task.

Thus, I am disinclined to adopt the stance of some philosophers of science, such as Nancy Cartwright, who tend to reply to questions of the sort 'How do we know that it is not true that, *ceteris paribus*, I can fly?' by pointing out that we do *not* know this, and that we would have to do some testing. Cartwright has presented a

---

[111] Imre Lakatos points out that the question of demarcation is, as a matter of historical fact, a question of life and death; see Lakatos, I. (1974). "Science and Pseudoscience" *Conceptus* **8**: 5-9. What question could be more important than a question of life and death?

sustained argument for the reality and *measurability* of capacities,[112] as well as for their seminal role in scientific theorizing. She urges us, for instance, to account for general causal claims such as ‘aspirins relieve headaches’ in terms of ascriptions of capacities realistically construed, rather than in an empiricist fashion in terms of the observed regularities involving aspirin. Closer to our subject matter, she acknowledges that capacity-ascriptions are not always unproblematic: ‘What the capacities of individuals are is another, very complex, matter. For instance, *must the relevant conditions for the exercise of the capacity be at least physically accessible to the individual before we are willing to ascribe the capacity to it?* These are questions I will have nothing to say about.’ (Cartwright 1989, p. 141; my emphasis). If philosophers of science want to deflect Kripke’s sceptical attack on disposition-ascriptions, it is precisely this type of question concerning the attribution of capacities to individuals, whose conditions of exercise are permanently inaccessible to those individuals, that needs answering. Cartwright, when pressed on the question how we are supposed to *empirically* test whether someone has acquired the capacity to add rather than to “quadd”—where the relevant *quus* function is defined as deviant in the higher reaches of the natural numbers—replies that Kripke’s skepticism, in this sort of context, is just ‘… *a version of the problem of induction and so not peculiar to capacities*. Do physical objects have the inertial and gravitational capacities we would naturally ascribe on the basis of Newtonian mechanics? No, not if we think relativity supersedes Newtonian mechanics. (I am making an analogy here between the “higher reaches of the natural numbers” and “very high velocities”).’[113]

Cartwright thus holds that the question whether ordinary cognitive agents endowed with infinite memory would be adding or quadding is analogous to the question whether ordinary massive bodies travelling near the speed of light would be obeying Newtonian or relativistic mechanics—and hence that it is an empirical question (albeit somewhat theoretically tainted, for simple induction by enumeration gives us the wrong answer). Similarly, Cartwright believes that the problem how to rule out the claim that *ceteris paribus,* I can fly, is not a philosophical, but an empirical one: ‘I don’t see in you any characteristics of the kind that my background knowledge associates with the capacity to fly. So I don’t see any ground for saying you can. Of course, perhaps you can—maybe there are properties I don’t know about that you

[112] Cartwright *Nature's Capacities and their Measurement*. Cartwright has her own definition of dispositions, and what distinguishes them from capacities (see *infra*). But this sort of difference is too fine-grained to play any role at the present stage of our discussion.
[113] Personal communication.

have that carry this capacity or maybe individuals can just have capacities that are not guaranteed by lawful connections with properties. But to have evidence you can fly, I take it, we need evidence for one of these. But similarly, to announce you can't, we equally need evidence. I think we do have evidence in our observations that this is not the kind of matter about which we see much variation among humans.'[114]

Cartwright's staunchly practical, anti-philosophical, approach to the problem does not, in my view, entirely do justice to the nature of our difficulty: true, the hypothesis 'I can fly' is likely to be amenable to empirical elimination, by searching for characteristics generally associated with the capacity to fly in various species; if the hypothesis is to come out true, I ought to possess at least some of these, notwithstanding the fact that I am human. For instance, the having of a large pair of lightweight wings, and enormous muscles to operate them, would come in handy. Given that I do not currently display any such property, it is quite safe to say that I can't fly. However, taking into account our discussion so far, the relevant question is of course whether it is the case that 'CP(I can fly)'—and generally, whether 'CP(humans can fly)'—with a *ceteris absentibus* reading of CP. Here, Cartwright's analogy with the case of Newtonian vs. relativistic physics is telling. She is correct in so far as it makes sense to consider the question whether cognitive agents with infinite memories confronted with huge addition problems behave in the same way as normal agents confronted with small problems, as comparable to the question whether bodies moving at the speed of light behave in the same way as bodies moving at much slower speeds. After all, in both cases our inductions ranging over the "normal" conditions fail to apply in the "abnormal" ones, and we need to take recourse to other considerations. But here is the crux: just as the equally practically-minded Fodor, Cartwright declines to concern herself with the question why *some* capacity-ascriptions to objects under certain "abnormal" conditions, such as infinite memory or near-light velocity, are both scientifically interesting and justifiable, whereas others are not. We may quite safely assume that her answer to this worry is similar to Fodor's, namely that this latter question, too, is ultimately for science to decide.

Yet, science does not seem to want to decide. Science does not explicitly tell us why it does not take certain seriously hypotheses, and in particular, why it would consider certain "abnormal" or ideal conditions just "too abnormal", or too ideal. The claim that there are ideal conditions under which any particular can do anything, is a metaphysical claim accepted by many philosophers, who, it must be said, are used to

[114] Personal communication.

abstruse possible-worlds arguments.[115] It is, as such, not in the purview of science. However, in order to prove that CP-clauses are not vacuously true, we need to prove precisely that either it is *not* true that there are ideal conditions under which any particular can do anything, or that CP-clauses do not in fact refer to *all* possible sorts of conditions, i.e. that some such conditions are outside their scope. A criterion for the acceptability of certain ideal conditions that merely adverts to *de facto* scientific practice, given science's lack of concern for metaphysics (in particular for the truth or falsity of the above claim), is thus not likely to establish that CP-clauses are not vacuous. The question, in Cartwright's own words, is not 'whether the relevant conditions for the exercise of a given capacity be at least physically accessible to the individual before we are willing to ascribe the capacity to it', for it seems clear that in the case of some capacities at least they do not have to be so accessible—it is, rather, *which* physically inaccessible conditions for the exercise of a capacity it is reasonable to accept before ascribing the capacity. Actual scientific practice tells us which such conditions it takes into consideration and which it ignores in actual ascriptions, but it does not tell us the grounds for its choice. Abstruse CP-ascriptions based on exotic conditions are generally *ignored* as opposed to falsified, which means, I suppose, that the possibility of their being (vacuously) true is not ruled out. Cartwright herself refuses, consistently with her position, to say that it is *false* that CP(I can fly), for although we do not have any evidence in favour, we also have, prior to examination, no evidence to the contrary. Clearly, in cases where no amount of examination is going to yield evidence one way or the other, because we are dealing with one of the more farfetched kinds of ascription, simple deference to actual practice will not solve our problem.

Greg Ray takes, in a related context,[116] a sweeping way out, and advances the following:

> … the only possible circumstances which can come about or be brought about—and hence the only circumstances which are candidate settings for

[115] I have cited Mumford and Lewis above. Lewis, in particular, is known for routinely employing considerations involving possible worlds in which the physical laws valid in this world do not hold, in order to solve specific philosophical problems relevant to this world, such as e.g. causation. Scientists, on the other hand, usually refrain from statements about such possible worlds. They generally dislike the sort of modal properties the philosopher is interested in (see Sec. 3.3).

[116] Ray, G. (1997). "Fodor and the Inscrutability Problem" *Mind and Language* **12**(3-4): 475-489; Ray discusses Fodor's appeal to inferential dispositions of agents in his attempt to solve the problem of explaining in a naturalistic way why 'rabbit' refers to rabbits rather than undetached proper rabbit parts. (see Fodor, J. A. (1993). *The Elm and the Expert: Mentalese and Its Semantics*, Cambridge, MIT Pr)

> scientific observations—are physically (or nomologically) possible circumstances. … physical dispositions of physical agents are only to be made sense of within the realm of physical possibility. *It is simply unclear what it would mean to say that a physical agent was disposed to apply a predicate in a physically impossible circumstance.* We are interested in the agent's actual dispositions, not counterfactual dispositions a speaker might have in a physically impossible world (Ray 1997, p. 480; my emphasis).

Sweeping, and simple. Ray comes close here to Schiffer's position, according to which we need to make sense of CP-laws (if we can) by consistently remaining in the realm of the physically possible. Again, the problem with this is that it would seem to rule out many *prima facie* legitimate disposition ascriptions to objects under idealized conditions.

This brings us to the close of this Section as well as the Chapter. The Section, I take it, has established that in so far as questions and problems of this kind are philosophical *par excellence*, science alone will not answer them for us. Philosophy's labour in this area, hence, is not lost. Our look at the respective achievements and failures of Fodor 1991 and Pietroski and Rey 1995 has uncovered a possible avenue of progress. This avenue points towards the necessity of a better analysis of the relationship between the attribution of non-manifesting dispositions, and underlying idealizational assumptions (or between the admission of CP-laws and the obtaining of physically impossible completers). These assumptions are idealizational, because they amount to implicit hypotheses as to which permanently, and perhaps necessarily, non-actual conditions would need to obtain for a given non-manifesting disposition to manifest itself. As such, ascriptions of at least some non-manifesting dispositions are comparable to explicit or implicit non-deductive theoretical inferences, inferences governed by the same sort of constraints and factors as influence our inductive inferences.

Chapter 2 has shown how a realist response to Kripke's paradox involves appeal to inhibited dispositions, finkish dispositions, conditional facts, or competences. These entities all share one important feature, I have argued, namely that we find out about them and postulate their existence on the basis of observations of manifesting or uninhibited dispositions, actual facts, or performance. In other words, inhibited dispositions, conditional facts, and competences, are related sorts of entities in so far as they are intrinsically unobservable entities whose existence needs to be inferred from the observed. The fact that the realist response to the Sceptic involves appeal to some theoretical entity or other, ought not be surprising: the “realist turn” in the theory of dispositions followed on the heels of the mentalist turn in the philosophy of mind, a central feature of the former as well as of the latter being the jettisoning of all behaviourist scruples with regard to the legitimacy of inferences from the observable to the unobservable. Just as there is no reason, for a realist, to suppose that merely because unobservable mental states are usually inferred on the basis of observable behaviour, we should consider the reality of the mental as more doubtful than the reality of behaviour, there is no reason, for a realist, to be sceptical about dispositions merely because all we ever observe are their manifestations. Realism about dispositions, we have found, is precisely the claim that in the absence of defeaters, even deeply buried dispositions or competences *would* cross the line from the unobservable to the observable and make themselves known to us through uniquely identifying manifesting dispositions, performances, or actual facts. For example, if I were not limited by my finite cognitive powers, I could compute the entire addition table. As things stand these sorts of dispositions or competences do *not* cross the line, of course, and often necessarily so, with the consequence that their existence must necessarily be inferred mediately from their observable counterparts that, unfortunately, severely underdetermine them. But the realist view is precisely that the presence of defeaters, albeit permanent, does not justify an anti-realist stance towards these entities. As we have seen in Section 2.4.2, the only realist who is moderately explicit about this in our context is Millikan, who acknowledges the obvious, namely that all evidence we could ever gather would not rule out quus-like hypotheses, and that the plus-hypothesis is preferred solely on explanatory grounds. What Millikan does not say is that the sort of inference to the best explanation that allows us to uniquely determine competence on the basis of performance—to establish the presence of an inhibited disposition on the basis of manifesting ones, or the obtaining of a conditional fact on the basis of actual facts, etc.—is an inference that contains a substantial element of *idealization*. For, as shown in Figure 5 (Sec. 2.4), the transition from the entities on the right to those on the left requires abstracting away from interfering factors, namely those causal influences on competences, inhibited dispositions, etc., that eventually “reduce” them to the actual finite performances, manifestions, facts, etc., we can actually observe. The elimination of causal influences, however, and the consideration of causally isolated systems, is the main characteristic of idealization in science.

Section 2.5 explored the ramifications of this observation: a solution of Kripke’s paradox demands, and the various realist solutions of it attempt to provide, a type of fact that is *theoretical* through and through, obtained through an inferential process that involves a substantial amount of idealization. Fodor 1990 deserves credit for making this entirely explicit when he briefly remarks on Kripke's paradox and offers his own solution. There is nothing wrong, Fodor suggested, in approaching our problem in terms of a psychological law ascribing, *ceteris paribus*, the disposition to add to agents (2.5.1). We criticized this on the grounds that Fodor's epistemology of CP-laws insufficiently illuminates the role of idealization, which, on his account, occurs when we choose the appropriate 'completer' for a given *ceteris paribus*-law. Fodor's theory lays insufficient constraints on this choice, and an improvement of it is suggested. The last Section went on to adapt to Fodor's case a counterexample in the literature to the very similar theory of Pietroski and Rey 1995, and showed how the suggested amendment escapes it.

3. *Ceteris Paribus*-Laws, Dispositions, and Idealization

This Chapter further explores the lessons from Section 2.5, namely that a successful account of why the *plus*-hypothesis is superior to its *quus*-cousin needs to identify some general properties of "inadmissible" idealizations. In 3.1.1 we examine Stephen Mumford's suggestion that the idealizations concomitant with disposition-ascriptions are "fixed by the context" of that ascription, and argue that when the context is the scientific one, Mumfor's proposal is tantamount to Fodor's. Section 3.1.2 then examines the potential of a suggestion by Mott 1992, according to which *ceteris-paribus* laws are best understood as *implicit descriptions* of (the data obtained in) experiments, by applying it to disposition-ascriptions. If we take disposition-ascriptions as implicit descriptions of "data", both actual and expected, then the idealizational component in these ascriptions could be explained as stemming from the idealization inherent in all modelling of data, or in all plotting of a curve over a set of points. The subsequent sections are thus concerned with the twin-notions of curve-fitting and idealization, with Section 3.3.1 arguing that every act of curve-fitting is a form of idealization, and Section 3.3.2 examining whether well-known constraints on curve-fitting, such as close *approximation* of the data, yield constraints on idealization. Curve-fitting, however, is more than mere approximation, and so is idealization. At this point we have, it seems, reached an impasse: if disposition-ascriptions are acts of curve-fitting with a substantial element of idealization, and if the only operative constraints on that idealization are the usual constraints on curve-fitting, then dispositional solutions of Kripke's paradox must fail (because Kripke's paradox, just like Goodman's, is a curve-fitting paradox). This negative result concludes Section 3.3. The last Section of this thesis argues that disposition-ascriptions, such as the ascription of the disposition to add, are "inferences to the best idealization", and that being an act of curve-fitting, so is the grue-hypothesis.

3.1 Disposition-ascriptions as Ampliative Inference

The upshot of the preceding section was that a better approach to the question, What makes spurious disposition ascriptions spurious?, requires reflection on the class of admissible idealizations. If what is wrong with the claim that CP (humans swim faster than dolphins) is its reliance on assumptions about what humans could do under farfetched "optimised" conditions, then the most promising approach must be to define constraints on the set of conditions that can sensibly be called 'optimal' or 'ideal' with respect to the question of the truth of the relevant claim. This task is rendered rather delicate by the fact that some idealizations—i.e. representations of objects that attribute to them some specific non-actual properties for reasons of theoretical, mathematical, etc. tractability—routinely employed in science are legitimate, whereas others apparently similar in nature and structure, are less so. Infinitely rigid levers, infinitely elastic molecules, the average British family, are acceptable, but infinitely strong agents and infinite cognitive powers, are not. Unfortunately, we cannot straightforwardly declare that the former sort of idealization is *true*, whereas the other is *false*. As Ronald Laymon[117] reminds us, all idealizations are by their very nature *deliberate misdescriptions* of reality. The use of idealizations that are strictly speaking false when taken as *bona fide* representations of actuality must lead to incorrect observational consequences. This seems to have the awkward implication that a theory employing idealizations is neither confirmable nor disconfirmable:

> Let t represent some underlying or fundamental theory, i the idealizing assumptions made, and p some actually derivable prediction. Philosophical and scientific common sense has it that if p is found to be true there is confirmation, or at least the satisfaction of a necessary condition for confirmation, and if p is found to be false, disconfirmation. But the falsity of i blocks such inferences. Consider first disconfirmation. From the premises $t \,\&\, i \Rightarrow p$ and $\sim p$ all that follows is that $\sim(t \,\&\, i)$, which is equivalent to $\sim t \vee \sim i$ … But this conclusion follows directly from the falsity of i. Therefore, there is no need to engage in the expense of experimentation if the conclusion sought is simply that either theory or idealizations are false. A similar problem holds for confirmation. (Laymon 1998, sec. 1)

[117] Laymon, R. (1998). "Idealizations" *Routledge Encyclopedia of Philosophy Online (http://www.rep.routledge.com/)*. E. Craig, Routledge, Taylor & Francis Group.

It would seem from this that idealizations are not hypotheses about the way the world actually is, but rather conscious misrepresentations of it with a purpose quite distinct from that of truthful description. If that is the case, then to chose between them—and in particular to determine whether a given idealization is an appropriate one to make when ascribing dispositions—we need criteria other than truth or falsehood, confirmation or disconfirmation. The most immediate response to this problem, the view that the best idealization is that which *approximates* reality sufficiently closely to produce reasonably accurate predictions, will be discussed in Section 3.2. First, we take a look at an attempt to characterize non-quantitatively what it is to be an 'ideal' condition for the manifestation of a disposition.

3.1.1 Context-relative Disposition-ascriptions

Mumford 1998 develops a functionalist account of dispositions according to which dispositional properties occupy particular causal/functional roles, i.e. they causally mediate in a determinate way between stimulus and manifestation. The functionalist idea is a natural one to try out when accounting for dispositions: any dispositional property will produce certain causal consequences given certain antecedent conditions, contrary to categorical properties, which are supposed to do what they do independently of any conditions. If an object is triangular, then it is triangular come rain or shine—which does not mean that it cannot cease to be triangular, it just means that its triangularity is not conditional upon any particular conditions. Clearly, dispositional properties are not like that, they more resemble a function that yields a certain manifestation (value) given the right initial conditions (argument). Thus, Mumford holds that the dispositional property of being soluble in water, S, has the functional role

WATER IMMERSION ⇒ S ⇒ DISSOLUTION

Mumford intends his functionalism about dispositions to be a *realism* about dispositions, in the sense that he explicitly states that the functional role of dispositions does not entail the truth of *any* counterfactual conditional. This is a paradigmatically realist move: recall Fodor, who held that idealized claims and theories can be true without entailing any counterfactual, "except the one the theory itself says is true". Dispositional properties are allowed to occupy functional roles without *ever* actually having to enact them, as it were.

Although an ascription of dispositions does not, on the realist picture, entail any particular conditionals, Mumford is aware that we still use conditionals in order to *specify* causal/functional role. Thus, it is natural to ask whether the same sort of worries that are associated with the conditional analysis of dispositions, accrue in equal measure to the functionalist one. For instance, in the case of counterfactual conditionals the presence of possible interfering factors not explicitly excluded in the antecedent makes it impossible to *guarantee* that the consequent follows. We may expect assertions that a property or a particular actually has *this* functional role rather than another to be subject to a similar kind of indeterminacy. As Mumford puts it:

> How can we say that something has a function to do ϕ if there is always the possibility of some interfering factor that will prevent it from doing ϕ? … this problem infects realist accounts … as much as it does an empiricist account. The realist says that disposition ascriptions are ascriptions of real powers. This leaves unanswered the question, 'power to do what?' The problem of background conditions means that the realist cannot say what it is that a power is a power to do. (Mumford 1998, p. 88)

Mumford correctly notes that we cannot simply exclude all possible interfering conditions in 'a finite list appended to the conditional'—or as Fodor would say, in a full description of the completer. The by now familiar problem is that we cannot hold, on pain of vacuity, that the excluded conditions are all those which prevent the disposition manifestation.

Mumford's remedy is the notion of *ideal conditions*. What a disposition ascription claims, minimally, is that '… a particular can do something. Given that this condition is met by most things, however, something more will usually be meant. What is usually implied by a true disposition ascription is that there are background conditions, let us call these 'ideal conditions', in which such manifestations do follow from the stimulus' (Mumford 1998, p. 88). Now, as we have seen in our discussion of Fodor, and as Mumford himself acknowledges, the appeal to ideal conditions in itself is still prone to vacuity, given that there are ideal conditions in which any particular can manifest any reaction (just as there is a completer for any CP-law). The way to interpret 'ideal conditions' non-trivially is to recognize that ideal conditions are *context-relative*, says Mumford: 'To say something is soluble is to say it will dissolve, in liquid, in a context relative to the ascription. The ascription in the actual world is relative to actual world conditions. It is also relative to actual world conditions that can vaguely be understood as 'normal'' (Mumford 1998, p. 89).

Mumford 1998 uses the terms 'ideal' and 'normal' somewhat interchangeably, but ultimately prefers 'ideal'. This ought to seem rather puzzling: after all, the meaning of 'ideal' seems very different from that of 'normal'! We clearly idealize when we regard air resistance as zero, but we do not appear to be idealizing when we say that under certain temperatures and pressures, salt dissolves in water.[118] Mumford's reason for running together the two is that he takes 'ideal conditions' by definition to be those background conditions in which the manifestation event associated with a given disposition follows from the stimulus conditions associated with it. It so happens that for many or most of our everyday disposition ascriptions, such as 'salt is soluble in water', these ideal background conditions are also meant to be those average temperatures, pressures, gravitational and electrical fields, etc., that we are most familiar with and that we regard as "normal" (Otherwise, many or most of our everyday disposition ascriptions would never be confirmed by their corresponding manifestation events, which would make it mysterious how we come to make them in the first place). Thus, in Mumford's usage, ideal conditions are *not* what we usually associate with the term 'ideal', i.e. quasi-impossible conditions realizable only in the laboratory and under great effort, if at all. For some authors, 'ideal conditions' are by definition non-actual, physically impossible conditions. Not so Mumford. The ideal conditions associated with a disposition ascription are, to him, those conditions under which we *most often* make the relevant disposition ascription. The point is that disposition ascriptions are also "indexical", in the sense that what we intend is a function of the circumstances in which we make the ascription. For instance, if I say right now—under usual temperatures, pressures, gravitational fields, etc.—that Peter is courageous, I do not explicitly or implicitly commit myself to any claim about Peter in *recherché* counterfactual circumstances in which one of these factors takes an extreme value—or in general to *any* counterfactual circumstances that might be expected to affect his dispositional state of being courageous. Obviously, temperature, gravitation, etc., are more likely to directly affect Peter's general physiological functions, rather than have differential impact on one of his psychological states. It is easy to imagine counterfactual circumstances, however, that would have precisely such an effect: brain damage; chronic depression; his being imprisoned and physically and mentally tortured by secret agents serving a malevolent foreign power, etc.

[118] I am indebted to Donald Gillies for this observation.

By making an ascription in a certain context, I implicitly exclude a set of counterfactual possibilities, and *which* set this is, so the idea, is a function of the context. 'In making an appropriate and useful disposition ascription I am saying that, in ordinary conditions for the present context, if a particular antecedent is realized, a particular manifestation usually follows' (Mumford 1998, p. 89).[119] If the present context were different, what qualifies as ordinary condition for the disposition manifestation would be different, too. Indeed, we do make disposition ascriptions that are true only under exceptional circumstances, as in scientific theorizing about objects under extreme conditions, e.g. to objects entering black holes, or objects under extremely low temperatures. On Mumford's account, we can say that 'is absolutely elastic' is a permissible disposition ascription in the context of theorizing about the behaviour of ideal gases, i.e. in the context of a possible world containig ideal gases—but that it is rather "out of place" when talking about, say, the properties of tennis balls. There may of course be extremely outlandish conditions under which even ordinary tennis balls would bounce without loosing energy, but these conditions are ruled out as *irrelevant* by most contexts of ascription—for instance, a Wimbledon match. In that specific context, the ascription to a tennis ball of the disposition to bounce indefinitely would have, on Mumford's view, "no point". To have a point, any given disposition ascription needs to be made in the *right* context, a context which will licence the assumption, if necessary, of unusual background conditions: '… the exceptional conditions will be fixed by the context of the ascription. That these are unusual background conditions will have to be flagged for the disposition ascriptions to have a point' (Mumford 1998, p. 90).[120] He adds, 'Of course, even in

[119] The regularity with which ordinary conditions obtain precisely is what makes possible inductive inferences about dispositions. We have frequently witnessed the manifestation before, as it followed the realization of antecedent conditions of a similar kind, and expect it to occur again in this instance. Our induction over antecedents under ordinary conditions is also the reason why we expect an explanation if the antecedent occurred without subsequent manifestation. Some post-Carnapian empiricists, such as Essler, W. K. (1970). "An Inductive Solution of the Problem of Dispositional Predicates" *Ratio* **12**: 108-115, have concluded that the problem of dispositions was a rather straightforward one of induction over the stimulus-manifestation event pairs. This flagrantly overlooks what we have emphasized above, namely that the main problem for empiricists about dispositions comes from the realist challenge to say something satisfying about dispositions that do not manifest and hence give us nothing to run our inductions over (especially those dispositions, if any, that necessarily do not manifest).

[120] By arguing that what we usually associate as the ordinary conditions for a given disposition ascriptions are in fact a subset of its ideal conditions, Mumford democratizes 'ideal conditions' and liberates them from the high-energy particle physics laboratory, or other exceptional places. I shall attempt to give rather different reasons for thinking that even "ordinary" disposition ascriptions such as 'salt dissolves in water' or 'Peter is courageous' contain an idealizational component in exactly the same sense as, say, 'All free-falling objects on Earth accelerate downwards at a rate of 9.8 m/s^2.'

these cases, the scientist concedes that something may interfere and prevent the expected manifestation.'

Let's see what this means for our ascription to Don of the disposition to add: as a matter of fact, for most numbers, Don does *not* have the disposition to add them unless very exceptional conditions obtain. On Mumford's account, if we nevertheless ascribe this disposition to him, then we are making a statement in a particular (presumably scientific) context, such that this context "fixes" the quite unusual background conditions (infinite memory, etc.) under which the disposition manifests. In a *normal* context, the ascription would have "no point", it would even be misleading: 'If … ideal conditions were exceptional, relative to the context of ascription, then there would be little utility in making the disposition ascription. Whoever made such an ascription would be, if not strictly speaking a deceitful ascriber of dispositions, at least an uncooperative or misleading one' (Mumford 1998, p. 89). In the context of scientific theorizing it is acceptable, then, to make disposition claims for objects which display these dispositions only under exotic conditions C_i, for these count *not* as exceptional, given the context. Under everyday conditions, however, it would be misleading to make any disposition ascription that presupposes C_i, given that in the non-scientific context these are exceptional. So the appeal to ideal conditions can be non-trivial: we cannot, in a given context, invoke just about *any* type of ideal condition to warrant a disposition ascription. Spurious disposition ascriptions based on exotic "ideal" conditions, such as my claim that I have the ability to fly, are excluded because the relevant background conditions under which these dispositions would manifest themselves are not warranted by the context of the ascription. Spurious ascriptions are 'misleading' and 'uncooperative' if out of context, even though not 'strictly speaking deceitful', says Mumford, and this is why it is that such ascriptions, when made in ordinary situations, need to be "flagged" as ascriptions that assume unusual background conditions.[121]

Curiously, Mumford thereby seems to suggest that spurious ascriptions are not strictly speaking *false*—just as it is not strictly speaking false, but merely unhelpful, to reply to the query "Where is Peter?" by saying that he is either in the pub or in the library (when one knows that he is in the library). Mumford accepts that even the most farfetched disposition ascription will be true on the background of *some* set of

[121] Mumford's approach to disposition-ascriptions bears similarities to the pragmatic account of CP-law statements advocated by Glymour, C. (2002). "A Semantics and Methodology for *Ceteris Paribus* Hypotheses" *Erkenntnis* **57**: 395-405, in so far as both are described to be speech acts that are, in important respects, *indexical*, and hence vary in truth value from situation to situation.

ideal conditions, and merely points out that assumption of the latter will be somehow uncalled for in most situations. The vagueness of this sort of proposal is compounded by the fact, recognized by Mumford, that given any type of context and any disposition, we can never fully specify the ideal conditions for the manifestation of that disposition in that context. This is due to the same reason for which a conditional analysis of dispositions is impossible: no finite list of test or stimulus conditions for a disposition manifestation will be able to *guarantee* the manifestation, because a further interfering condition is always possible (Mumford 1998, p. 88). This is also why the "conditional conditional"

If ideal conditions C_i obtain, then if T(x), then R(x)

although it is, according to Mumford, "invoked" by disposition ascription D(x), could never be understood as a reductive analysis of D(x).[122]

Mumford's picture is problematic from two points of view, one semantic, one epistemological. Concerning the semantics, the worry is that Mumford's ideal conditions for a disposition are allowed to change from one (conversational?) context to another, although they are, in some sense, supposed to be part of the *meaning* of the dispositional predicate used in the ascription. This seems to have the consequence that our dispositional predicate can have different meanings in different contexts, but the dispositional *concept* we are dealing with remains the same (cf. Malzkorn 2000, p. 459). (This is significantly different from the case of true indexicals, such as "here" or "I", whose meaning and associated concept is the same across contexts, though its referents change). To this Mumford replies that, on his view, the dispositional predicate D is associated with just *one* dispositional concept as long as the ideal conditions C_i remain unspecified; however, different, more precise dispositional concepts become associated with it as soon as we begin to characterise C_i more fully (Mumford

[122] Cf. also Mumford, S. (2001). "Realism and the Conditional Analysis of Dispositions: A Reply to Malzkorn" *Philosophical Quarterly* **51**(204): 375-79, at pp. 375-76. One can understand Mumford's reticence, as a proclaimed realist about dispositions, to even consider the possibility that such 'conditional conditionals' provide an analysis of disposition ascriptions, or are entailed by them. For the refusal of the traditional (empiricist) conditional construal of disposition predicates is of course the central plank of his platform. Nevertheless, we are entitled to wonder why, if the ascription of D to x has no counterfactual implications whatsoever, such conditionals should be "invoked" by it in the first place (what kind of relation is that?). A philosopher with fewer realist scruples might very well be tempted to accept as a rough analysis of D(x) this conditional conditional, or conditional defined over ideal conditions, for the idealized status of C_i protects this sort of approach from the problems with necessarily unmanifesting dispositions that plagued Carnap and Goodman. Moreover, it establishes a link to idealized scientific laws, which are usually thought to have conditional form.

2001, p. 377). He could conceivably also argue that the concept deployed remains the same across all applications, never mind ideal conditions. The predicate ‘courageous’, for instance, admittedly has one meaning when applied to children in the playground, and another when applied to soldiers in the battlefield—in the sense that its criteria of application differ, just as the word ‘tall’ has a different application in Sweden than it has in Sicily—but it is not clear that this notwithstanding, it is not the same *concept* at work across all applications.

To follow the ramifications of this sort of debate would lead us far astray. The epistemological point shall exercise us more: it seems that we are not in a world where hypothesizing about what would happen if there were C_T-radiation makes much sense. The claim that, given C_T-radiation, turtles have extraordinary dispositions ought to count not merely as true but misleading and uncooperative. Such a hypothesis is, in Wolfgang Pauli’s phrase, “not even false”.[123] On the other hand, we do live in a world where it makes perfect sense to hypothesize about what would happen if molecules were absolutely elastic. The claim that given perfect elasticity, impermeability, etc., gas molecules in a container display a regular behaviour when we increase the temperature or decrease the volume of the container, ought to count not merely as useful and cooperative, but as *true*—and this in spite of the fact that perfectly elastic molecules, too, are nomologically impossible.

In view of our confrontation with the Sceptic, we need to be able to say not only when spurious disposition ascriptions are inappropriate to the conversational (or textual) context of ascription, but also when they are plain *false*, and when they are more than just false, when they wear their falsity on their sleeves. Clearly, farfetched ideal conditions invoked by spurious ascriptions are conditions we have no reason to consider as in any way relevant to the question of the truth or falsehood of a given disposition ascription to objects in *this* world. Thus we have no serious reason to take into account C_T-radiation when assessing a real turtle’s running capacity, and consequently no serious reason to ascribe the corresponding spurious disposition to the turtle. Of course, there is always the “theoretical possibility” of something like C_T-radiation, but it is a possibility not taken seriously by scientists. Note that the difficulty of showing what, precisely, should or should not be taken seriously when as-

[123] Liu, C. (1999). “Approximation, Idealization, and Laws of Nature” *Synthese* **118**(2): 229-256, p. 230, reports Wolfgang Pauli as having frequently used the expression “not even false” to refer to hypotheses so ill-formed that they should not even be considered candidates for approximate truth—in the way in which some idealizations are supposed to be approximately true—let alone candidates for factual truth. Such hypotheses are “*more* than false”, we may say, they wear their falsity on their sleeves, i.e. they are absurd.

cribing this-worldly dispositions is not restricted to typical philosopher's examples involving fantasy events such as turtles overtaking humans. The statement 'this month Don has learned how to add in school' ought to be able to come out as wither true or false, although it amounts to an ascription of the disposition to add, a disposition whose (full) manifestation requires conditions *prima facie* just as exotic as those invoked in the turtle case. The gist of the present discussion is precisely that our grasp of idealization and modalities is insufficiently developped to show that although the former is fantasy, the latter is not.

Mumford's distinction between contexts, and the context relativity of ideal conditions, provides little help here. What helps even less is that he sometimes seems to suggest that disposition-ascriptions are essentially "world-indexed", in other words, that their associated ideal conditions are nomologically possible in this world (in correspondence). This would wreak havoc with many scientific disposition ascriptions. Moreover, my statements about the arithmetical abilities of a student aren't particularly scientific in the first place (I am not engaged in theorizing about him, merely reporting his academic progress to my neighbour), nor are they particularly misleading. Even if it turned out that everyday disposition ascriptions are in some sense proto-scientific—like folk psychology in general, according to some authors—this would not solve the problem. Mumford's assertion that CP-disposition ascriptions that imply the assumption of extraordinary conditions can be warranted if made within the context of a successful scientific theory essentially mirrors Fodor's view that CP-disposition ascriptions are legitimate if they are warranted by a law that is a member of an appropriate network of true laws. Just as Fodor, Mumford is left with the task of specifying *how*, precisely, science—or speakers in the "scientific context"—goes about excluding spurious ideal conditions. What is it about the scientific context that de-trivialises disposition-ascriptions? Contrary to Fodor, Mumford does not address this question. [124] Thus, although he is aware that 'there are ideal condi-

[124] Wolfgang Malzkorn makes a similar criticism when he points out that the context-dependence of 'ideal conditions' threatens to make dispositional predicates trivially true. After all, for any dispositional predicate and any object, there are (as Mumford acknowledges) 'ideal conditions' such that the predicate can be applied to the object: a rose is not fragile at room temperature, i.e. under ordinary conditions of ascription, but it is fragile at –272°, a possible "scientific" context of ascription—and so are most other things. (Malzkorn "Realism, Functionalism and the Conditional Analysis of Dispositions", pp. 458-60). This would rob the predicate 'fragile' of its power to distinguish between objects, unless we are provided with further constraints on which conditions may, or may not, be invoked. A similar thought underlies my ascription to myself that I can fly: surely, under *some* extreme conditions of ascription, this might turn out to be true. All extreme conditions, in particular physically impossible ones, fall under the scientific context, and we therefore need further detail on how the scientific context de-trivialises disposition ascriptions. Mumford's reply to Malzkorn (Mumford "Realism and the

tions in which any particular can manifest any reaction', we find insufficient constraints in his theory on eligible ideal conditions. Again, the question is: why infinite memory and not infinite strength?

Mumford notes that his own earlier attempt (Mumford 1996a) to solve the problem of necessarily non-manifesting dispositions (e.g. Martin's electro-fink case) was unsatisfactory because by looking for restrictions on C_i, it amounted to adding a *ceteris paribus*-clause to the 'conditional conditional':

CP (If C_i, then if T(x), then R(x)) (Mumford 1996a, pp. 86-87)

Mumford takes it for granted that we have no way of non-trivially filling out the CP-clause, and hence assumes that this sort of proposal is inadequate (ibid.). He believes that by "sticking firmly with realism" and foregoing an analysis of functional role in terms of conditionals he circumvents the problem, with the doctrine of context-relativity taking the brunt of the burden of putting restrictions on C_i. But it is difficult to see how thus vaguely context-indexing a given disposition ascription yields a non-trivial interpretation of C_i. The central objection to the kind of approach sketched by Mumford is that even if some or most contexts of a disposition ascription flatly *ruled out* an appeal to certain extreme ideal conditions (as opposed to merely making it unwarranted or unhelpful), it would still be the case that *any* disposition ascription D(x) is true, *ceteris paribus*. This is because the truth of D(x) would simply be contingent upon having been uttered in the *right* context, warranting appeal to the *right* ideal conditions C_i. Even if that context is comparatively rare—in other words if it is rather uncommon that the *ceteris* are *paribus*—the mere availability of such a context suffices to make D(x) true. Mumford, we see, allows the CP-clause to enter through the back door. Holding that

(If C_i, then if T(x), then R(x))

is correct depending on the context is merely holding that there is a function CONT that takes contexts and disposition-ascriptions as arguments and yields truth values,

CONT(If C_i, then T(x), then R(x))=T

Conditional Analysis") fails to provide this detail. There he merely allows that 'fragile' may have been a bad example, and furnishes further examples, which, however, seem to fail for the same reasons.

This does not seem significantly different from

CP (If C_i, then if T(x), then R(x)),

insofar insofar as the CP-operator can *also* be read as a function, this time from conditions and propositions to truth-values. It suffices to interpret the relevant conditions as 'contexts of ascription' to establish the complete equivalence of the two proposals.[125]

If the relevant function remains unspecified—in other words, if we have no precise theory of how scientific context rules out infinite strength, C_T-radiation, etc.,—or alternatively, if we have no theory of CP-operators, then disposition-ascriptions will remain vacuously true.

3.1.2 Disposition-ascriptions as Curve-fitting

Peter Mott, who criticises Fodor for 'telling us nothing at all about CP-laws', suggests the following: explanation of experimental results being an important task of cognitive psychology, why not assume that CP-laws are in fact implicit *descriptions of experiments*, where 'experiment' is to be understood in a wide sense (Mott 1992). In Mott's proposed usage, the term applies both to what we usually associate with it, i.e. the controlled instrumental testing of a hypothesis in a laboratory, as well as such things as the throwing of a stone against a window. An experiment is, as Mott puts it, a two-stage procedure, in which we first set up the right conditions (if necessary) and start the experimental process, and then see what happens as a result. Just as the form of a law is usually thought to be conditional, the idea of conditionality is also implicit in our understanding of an experiment, says Mott, although this need not necessarily show up in the use of an 'if … then' clause: '… a small child says "I gonna step in puddle with sandals and get it all wet". … This is an experiment!' (Mott 1992,

[125] The air of equivalence is enhanced when we consider pragmatist accounts of CP-laws, such as Glymour "A Semantics and Methodology", that make much of the inherent *indexicality* of CP-claims: '*Ceteris paribus* generalizations, or claims that *normally, X*, are universal conditionals in which the antecedent is indexical and typically unexpressed; put another way, the qualifier "normally" acts logically as a propositional function variable whose value is somehow determined in each case. *Ceteris paribus* generalizuations are true for a case, a circumstance, a situation, if the appropriate value for the indexical antecedent for that case results in a true conditional...' (Op.cit., p. 400).

p. 341).[126] Every CP-law refers, according to Mott, to a specific procedure that is the 'underlying experiment' of the law. This yields a practical way for establishing whether a given conditional statement deserves the status of CP-law, namely by performing its underlying experiment (the antecedent of the law describes the preparation/performance of the experiment, and the consequent the result). 'No law without experiment', quips Mott, referring to Ian Hacking's influential defence of experimental science (Hacking 1983) as the inspiration for his approach.

Laws are *ceteris paribus* precisely because their underlying experiment sometimes fails, with the experiment refusing to produce the result stated in the consequent of the law. Mott elaborates:

> This, it should be emphasized, is not to say that the law is sometimes falsified and sometimes verified, still less that it is sometime true and sometimes false. The law is always true (*ceteris paribus*). Its *experiment* sometimes does not work. It fails to work for one of two broad reasons: insufficient skill on the part of the experimenter, or random interferences from the world. (Mott 1992, p. 342)

Assuming that we have independent criteria for 'sufficient experimental skill', Mott proposes a rough yardstick for distinguishing mere interference from genuine falsifying instances, as follows: to count as simple glitch that does not discredit the law, a failure of the experiment needs to be attributable to 'an irregular range of chance interfering factors' that are sufficiently few (ibid.). The way to achieve this attribution is again through practical experimenting, namely by attempting to reproduce the result of the experiment *without* the aberration. Only if such attempts, pursued with sufficient experimental skill, lead to a systematical reproduction of the 'aberration' or 'anomaly' as well as of the normal effect, ought we reject the law.

Mott uses this account to argue that if laws always are descriptions of underlying experiments, then Schiffer 1991 is at least partly correct, in the sense that Folk Psychology does not actually provide us with CP-laws whose antecedents contain reference to intentional states (e.g. laws connecting beliefs and desires with behaviour). All we can hope for are psychological laws with physical states in their antecedents (e.g. laws connecting a physical stimulus with an intentional state). The asymmetry is due to the fact that whereas we can construct an experiment for the latter

[126] Mott quotes from Manktelow, K. I. (1990). *Inference and Understanding: A Philosophical and Psychological Perspective*, New York, Routledge, here. The example is intended to illustrate that the preparation and performance of an experiment need not always be entirely distinct from each other.

kind of laws by manipulating the relevant physical conditions, we cannot set up experiments for the former, because we simply cannot manipulate people's beliefs and desires in a comparable way (Mott 1992, p. 343). We shall not scrutinise the merits of this claim here—some may wish to argue that, on the contrary, there are quite controled and precise experimental setups for manipulating people's beliefs comparable to physical experiments—for it is beyond the scope of our present concerns. We will, however, examine Mott's way of distinguishing between mere 'random aberration' and the conclusive falsification of a CP-law.

At first sight, Mott seems to offer nothing new on this front: CP-laws are conditional statements that are true, unless something interferes. However, his way of separating the good interference, which does not falsifiy the law, from the bad, is different, for it is not based on either requiring occurrence in related law statements (Fodor), or explanatory virtues (Pietroski and Rey). Mott's is a purely operational criterion inspired by experimental practice: to find out whether your anomalous data is due to random noise, rather than an indicator of stable, systematic interfering factors, repeat the experiment and try to establish a pattern. If no such pattern materializes, you may consider the relevant exceptions to the CP-law as instances of precisely those tolerated exceptions that make the law *ceteris paribus* in the first place. Unfortunately, Mott's operational criterion fails to take on board the complexities of the relation between observational data and CP-laws, on at least two counts. Firstly, the idea that CP-laws are implicit descriptions of experiments does not generalise easily beyond a narrow range of experimental laws in cognitive psychology. There are scores of laws in almost every science for which there can be no actually executable procedure that constitutes their underlying experiment, because they are either defined over ideal conditions, or because they are higher-order theoretical principles designed to unify existing, more phenomenological/experimental, laws. Given Mott's emphasis on the priority of experimental science, we can safely presume that he did not mean to include *impossible* experiments (i.e. thought experiments) into the category of 'underlying experiment'.

Mott may quite legitimately retort that his criterion was not in fact intended to generalise beyond CP-laws in experimental cognitive psychology. More troublesome for his account, then, is the fact that what we actually measure, in *any* experiment, almost never coincides exactly with what the relevant law says we should measure—and we do not expect that to happen, either. For most laws we fully expect a rather constant pattern of interference, so much so that if in one instance our measurements happened to coincide *exactly* with the predictions, we would feel compelled to investigate what had "gone wrong". Many CP-laws, among which the philosophically interesting ones, do not encounter an 'irregular range of chance interfering factors' that are sufficiently few, but rather a regular range of quite constant interfering factors, such as e.g. friction in mechanics, or irrationality in economics. Cognitive psychology is no exception, where distraction, limited memory, etc. are permanently interfering. Moreover, in many cases we do not have even a rough grasp of the kinds of interfering factors at work, and cannot estimate their number. For instance, we do not know how many factors, precisely, interfere with a molecule's being perfectly elastic: on the face of it, it is the electrical attraction between molecules—but does this mean that the fact that electrical force is one of the fundamental forces of nature also counts as an "interfering factor" with respect to the law in question? If yes, ought whatever it is that makes the non-existence of electrical force impossible (in this world), also count as interference, and so on? (Pietroski and Rey, reflecting on a similar point think not). *Ditto* for my brain's ability to manipulate large numbers, or its capacity to remain permanently focused on the task at hand. In Fodor's terminology, many CP-laws encounter absolute, and not only mere exceptions. This is because, as we have already pointed out, these CP-laws are in fact *ceteris absentibus*-laws, they typically claim that if certain often vaguely specified interfering factors were absent , although they are in fact permanently present, then if A then B (cf. Joseph 1980). Mott's injunction to attempt to repeat the experiment without aberration would, in all those cases where this is physically impossible, lead us to reject it.

Nevertheless, all is not to be disregarded in Mott 1992. What shall interest us most is the simple idea that CP-laws are nothing over and above the interpretation (description) of the results of experiments. The suggestion in itself is, of course, nothing new—it is in the spirit of a broadly anti-realist philosophy of science that views theories as mere *instruments* for the control and prediction of nature. We shall apply a double transformation to this suggestion to let it do work in our context. Firstly, we need to enlarge the category 'underlying experiment' to include physically impossible experiments (i.e. thought experiments), in order to cover the cases that concern us. For example, the statement that *ceteris paribus*, the pressure (P) and volume (V) of a gas in a container varies with the amount of gas molecules (N) and their temperature (T), shall be understood as the description of a *physically impossible* experiment in which we confine a known quantity of an ideal gas in an impermeable container, increase T, and then measure P. Secondly, we use the fact that disposition ascriptions, in Mumford's words, "invoke" conditional CP-claims of the form

CP (C → (S(x) → M(x)),

and apply Mott's idea to those disposition ascriptions, rather than to the CP-laws themselves. Thus the theory to be tested presently is that a disposition ascription 'object X has disposition D' is, in fact, an implicit description of actual or possible experiments in which we observe X's behaviour after it has been subjected to the (disposition-specific) stimulus or test conditions S, under (disposition-specific) background conditions C. For example, the disposition ascription 'Salt is water soluble' is the implicit description of certain experiments with batches of NaCl crystals, whose results can be graphically represented as follows.

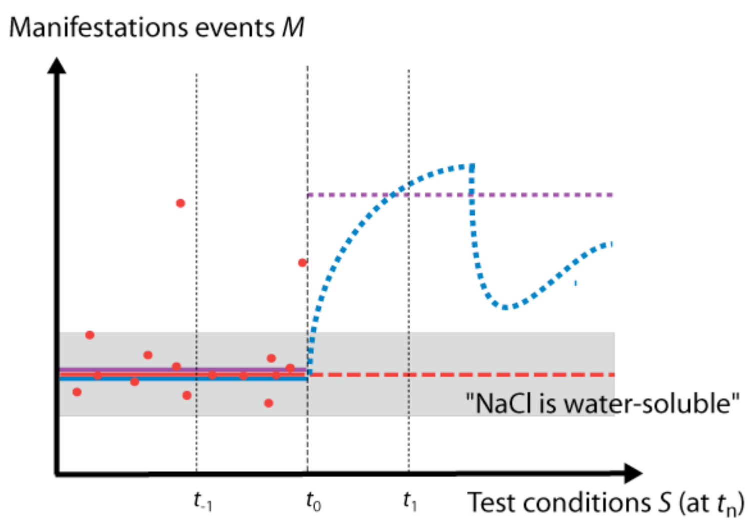

Fig. 8

We may conceive of each red dot as an event, observed at t_n, which is the result of subjecting a given quantity of NaCl to test conditions S at t_n. A test that has been duplicated by exactly reproducing conditions *S* counts as a new test, because its results have been observed at a later time. The grey rectangle indicates the range and kind of manifestation events we (conventionally) interpret as manifestations of solubility. Solubility, as most other properties, is not an all-or-nothing affair. In the present case, the range might be the result of a double constraint, one quantitative, and the other temporal: a substance which leaves a large residue, and/or takes several years to dissolve, does not count as water-soluble. The types of constraints defining the range of eligible manifestations will often be somewhat vague, and vary wildly from disposition to disposition. The two outlying dots beyond the grey zone are anomalies, cases of individual batches of NaCl crystals not dissolving in water within the prescribed time limit, or not fully enough, or both. These we account for as cases in which, either, the disposition-specific background conditions C must have failed to obtain, or further interfering factors that escaped the attention of the experimenter entered the fray and prevented the salt from dissolving. The data points indicate our experience to date (t_0), and the red dotted line represents our expectations (predictions), based on our belief in the correctness of the disposition ascription, concerning the mean reaction of batches of NaCl subjected to conditions S in the future.

To say 'Salt is water-soluble", on this story, is thus the very same act as fitting a curve onto actual experimental data, and thereby making a claim about possible experimental data.[127] The function generating the curve maps a set of certain stimulus conditions at certain times, $\{S,t\}$, onto the set of manifestations, $\{M\}$, yielding the set $\{\{S,t\}, M\}$. The disposition ascription amounts to a retrodiction, specifying the behaviour of NaCl patches in the past, had they been subjected to the test at t_{-1}, as well as a prediction of possible results of tests at t_1 in the future. Moreover, in virtue of being an act of curve-fitting, uttering 'Salt is water-soluble' is *eo ipso* excluding incompatible possibilities. For, Salt might be ascribed *other* kinds of disposition, whose existence, although consistent with the actual data, implies conflicting retrodictions and predictions. Thus, the purple line in Figure 5 represents a grue-type disposition ascription to salt, whereby batches of NaCl have the disposition to dissolve if $t < t_0$, and to not dissolve if $t > t_0$. We may imagine ascriptions of varying complexity, e.g. the blue line, infinitely many of them. An interesting consequence of the present view is that our reasons for saying that salt is soluble, rather than possessing any other, more complex disposition, are *exactly the same as* the reasons we deploy in standard curve-fitting procedures, such as the concern to find an acceptable compromise between empirical adequacy ("goodness-of-fit"), and the simplest explanation (see the following Section).

Non-scientific cases, such as folk-psychological disposition ascriptions, ought to be no exception, on our picture. Thus, take the innocuous disposition ascription 'Peter is courageous':

[127] True, one might object that saying that 'salt is water soluble' just predicts that a red dot will lie in the grey area, which does not seem to involve *curve* fitting properly speaking. However, this is simply a device intended to account for the fact that most dispositional predicates are such that their generic test conditions encompass a rather wide range of conditions, sometimes vaguely specified.

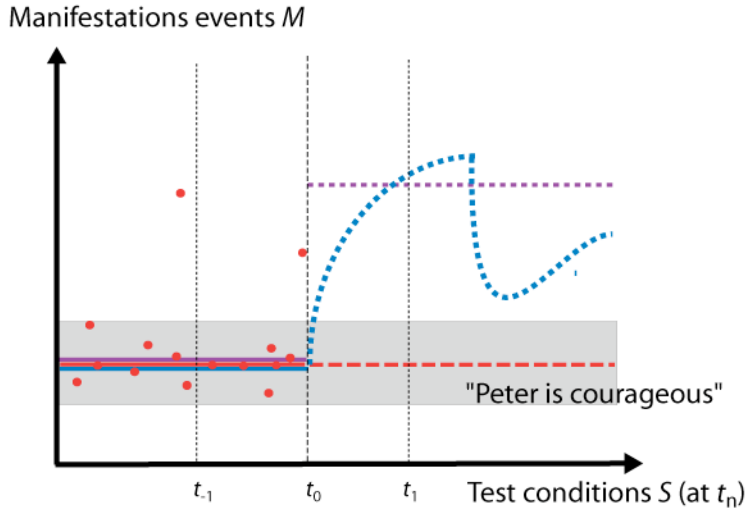

Fig. 9

Here, each dot represents an observation of Peter in conditions S, t behaving in a certain way, where S are assumed to be "courage-relevant" circumstances, and hence to constitute test conditions for his courage. This is somewhat trickier than in the *NaCl* case: Peter's standing in front of the fridge at 8:00am, and proceeding to open it, is neither a test condition, nor a manifestation of courage; his standing in front of a burning house with a helpless child inside, and his proceeding to enter the house to rescue the child, on the other hand, qualifies as a member of $\{S,t\}$ and $\{M\}$, respectively. Again the grey rectangle in the graph represents the range of behaviour $\{M\}$ that we conventionally characterise as "courageous". Ideally, every observation of Peter in conditions S will either confirm or disconfirm that he is courageous, depending on whether his action falls within the specified range. Obviously, the picture cannot be as tidy as in our previous graph. Psychological cases require us to introduce some latitude concerning what the data points are allowed to represent (as in Figure 3, *supra*). After all we may ascribe courage to Peter not on the basis of direct observation of behaviour, or any other kind of event involving him, but rather trust the testimonials of others, or historical sources. We may even do it on the basis of a hunch, of indirect and inconclusive clues, which we are unable to make explicit. In fact, 'courageous' is one of those interesting psychological predicates denoting a psychological state whose self-ascription is rather prone to error and probably not any more precise than its ascription by others. No one really knows reliably if he or she is courageous prior to having experienced test conditions for courage.[128]

What matters presently is that there is a way of looking at disposition ascriptions which reveals that they have the general form of curve-fitting, albeit curve-fitting in which the "data" is not necessarily numeric, and where the "curve" fitted is not really graphic, either. For what is being fitted is the function $\{\{S,t\}, M\}$—or, if you prefer, a *meaning* for the predicate 'courageous'—which, in psychological cases, can only at the price of some akwardness be represented graphically. Nevertheless, the comparison of disposition-ascriptions and curve-fitting is intended as more than merely a loose analogy. Curve-fitting and disposition ascriptions have something important in common, in so far as both are *acts in which we unify a number of potentially disparate observations under one header*. We have different options what to call these acts. Psychologists will say that they are acts of categorization, whether conscious or unconscious, in which case they are categorical perception. Some philosophers might prefer to say that they are acts of recognizing the instantiation of a universal in a particular. We liken them to curve-fitting here (and sometimes call them that), for reasons which shall become clear. Granted, it is more natural to speak of 'curves' and 'functions' when the data is discrete and quantitative, than when what we are talking about is kinds of behaviour under certain conditions—one might even prefer to speak of "description-fitting" instead of curve-fitting in the latter case: does the description 'courageous' fit this behaviour? However, I shall not inflate terminology and stick to the one term curve-fitting; for the present suggestion is not that disposition ascriptions *literally* are acts of curve-fitting (this would be absurd), but rather that both disposition ascription and curve-fitting involves doing the same sort of thing.

What interests us is thus that both curve-fitting and disposition ascriptions share important properties, possibly enough to meaningfully classify them as (based on) the same kind of inference. Both are clearly inferences of the non-deductive, ampliative kind: the ascription of courage to Peter is an ampliative inference which constitutes a retrodiction as well as a prediction of Peter's possible behaviour, making implicit claims about what he would have done had he been subjected, at a certain time, to test-conditions for courage. Knowing that the ascription is true is therefore knowing more than the sum of our true observations of Peter. Analogously to the sta-

[128] The same, I suspect notwithstanding the mentalist climate in current philosophy of mind, might be true of more psychological attributes than we think. As Wittgenstein puts it 'love is something that is put to the test'.

tistical curve-fitting case, we may make certain observations of Peter outside of the grey area which appear to contradict the ascription, and we may chose to ignore them as anomalous—i.e. assume that those cases were either "flukes", or unknown interfering factors were at work. (E.g. Peter's having a severe phobia of fire, which paralysed him and thus inhibited his otherwise courageous reflexes to jump into the burning house and save the child). The decision to ignore observations that are considered not "characteristic" is isomorphic with the decision to ignore outlying data while plotting a curve over data-points. It is governed by the same sort of criteria as curve-fitting in general.

For example, in theorising about—or merely non-scientifically reflecting upon—Peter's character traits, we prefer the simpler hypotheses over those that account for each and every observation we have made, i.e. those hypotheses that achieve a very high degree of fit at the cost of complexity. Moreover, we like those hypotheses that better unify our pre-existing data, and better explain it—e.g. we prefer to ascribe attributes to Peter that better fit with the *other* things we know about him, and we want to ascribe attributes that *best explain* his observed behaviour. Just like in curve-fitting, we deal with anomalous observations by drawing on our background knowledge of the object under study, and just like in curve-fitting, we want to obtain a description that *best explains* the data. It is no good to ascribe the attribute 'is courageous OR likes Mozart' to Peter on the basis of observing him running into the burning house if we have no independent reasons for thinking that his running into the burning house might also somehow be indicative of his musical taste, just like it is no good to choose an overly-complicated function for our curve if we have no independent reason to suspect that the anomalous parts of our data are due to a highly complex interaction of causal factors. Finally, just like in curve-fitting, we simply discard the obviously absurd possibilities, as e.g. the grue-like hypothesis that Peter is courageous if observed before t_0, and not courageous if observed after t_0.

It may be objected that not all of so-called folk-psychology consists of disposition-ascriptions so construed, and, indeed, that *most* of folk-psychology cannot be so construed. Some may find generally objectionable any assimilation of everyday disposition-ascriptions to what is essentially a practice confined to laboratories. It is true that, as Daniel Dennett puts it, folk psychology is not exclusively about making bets what others will do (Dennett 1991, p. 29), and hence not about prediction. Curve-fitting on the other hand is, as I have stressed, a form of prediction as well as retrodiction—in other words, it is a bet about what someone or something, given certain conditions, *would* have done. However, one of the most puzzling features of folk psychology as we practice it is precisely the great power it gives us to reliably interpret the actions of other people, understand them, to put ourselves in their position, etc. These capacities, Dennett notes, crucially depend on our power to actually *predict* others, even if we rarely explicitly exercise it: '... at the heart of all these [capacities] is the enormous predictve leverage of folk psychology. Without its predictive power, we could have no interpersonal projects or relations at all; human activity would be just so much Brownian motion; we would be baffling ciphers to each other and to ourselves—we could not even conceptualize our own flailings' (Dennett 1991, Ibid.). Whether we consciously engage in such predictions or not, they must be at work at some level of our cognitive relation with others. The success of prediction, and hence of folk-psychology in general, however, depends on there being some pattern, or regularity, to exploit—without patterns *nothing* would be predictable, as Dennet likes to say. In our terminology this translates to: the success of disposition ascriptions depends on the success of a curve-fitting-like inference. We need to fit the *right* curve, in other words choose the right description and detect the *right* kind of regularity. The great difficulty is, of course, as Goodman put it, that "regularities are where you find them", that everything is multiply redescribable.

3.2 Curve-Fitting and Idealization

Contrary to appearances, the topic of curve-fitting is not at all removed from our previous concerns about idealization and the correct constraints on ideal initial conditions. Curve-fitting is, as we shall see, a form of idealization, and its importance for our problem consists precisely in the fact that it represents a conspicuous case of idealization more easily amenable to analysis, perhaps, than idealization in its other guises.

3.2.1 Curve-fitting as Idealization

The fitting of a curve on a set of discrete measurement points can be described as a form of *idealization*. The reason for this is simple. Suppose you make a series of simultaneous measurements of two quantities x and y, and that you obtain the following results:

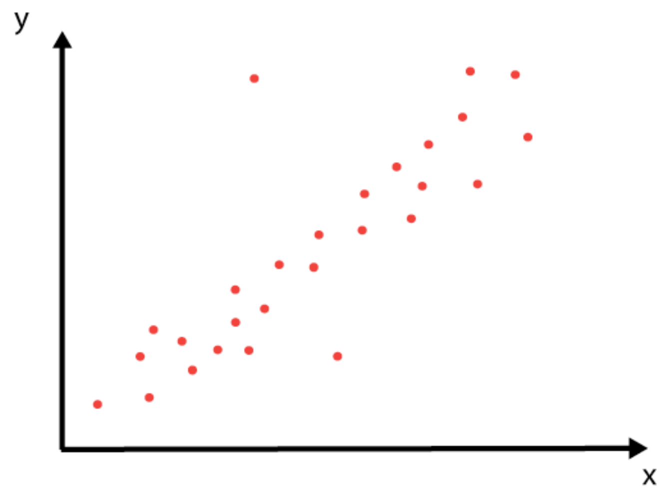

Fig. 10

A short look at the data will suggest to you that something significant is going on: x and y seem to stand in a functional relation to each other, for if both x and y were in fact randomly distributed, the probability of obtaining a result such as this one would be small, especially as the number of measurements increases. Also, you will very likely come to think that the functional relation between x and y is linear and continuous, of the form

$$y = ax + b,$$

In attempting to find that relation, you will thus try to plot a straight, continuous line, over the data. The question then becomes: which such line best "fits" the data? According to the 'method of least squares', we obtain the best-fitting straight line,

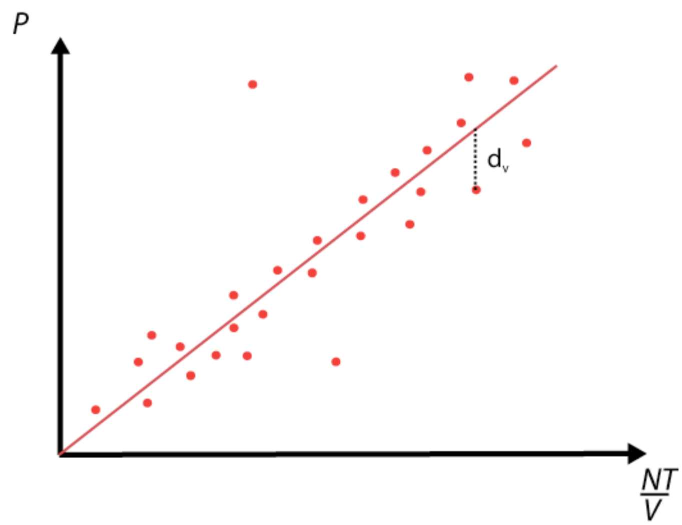

Fig. 11

by minimizing the sum of the squares of the vertical distances, d_v^2, between each data point and the curve (alternatively, by minimizing the sum of the squares of the horizontal distances, d_h^2).[129]

Note that the particular straight line we have obtained through the least squares method does not actually go through all of the points; in fact, in our example, it goes through *none* of them. If we needed a curve that went through all or most of the points in the data—for example, because we have good reasons to assume that the measured quantities stand in a more complex, non-linear relationship, or because we have high confidence in the accuracy of each measurement—we would have to abandon the assumption that the function we are after is a linear and continuous one, and would perhaps attempt to fit a rather more complex polynomial. Although the method of least squares, or 'linear regression', is a standard statistical solution for obtaining the closest curve once we have made up our mind concerning the general shape of the function we are looking for, it does not tell us how to initially choose the kind of function we need. If we choose to fit a more complex function, it does not tell us which among the infinitely many alternatives achieves the closest fit. In fact, when it comes to non-linear regression, there is no universally agreed best method. Given any finite set of points there will of course always be infinitely many curves that go through all of them, and we might have extraneous, non-"data-driven" reasons for preferring one of them over the others. Curve-fitting, so the practitioners say, is as

[129] For an accessible and philosophically informed discussion of curve-fitting, see Forster, M. and E. Sober (1994). "How to Tell when Simpler, More Unified, or Less "Ad Hoc" Theories Will Provide More Accurate Predictions" *British Journal for the Philosophy of Science* **45**(1): 1-35.

much art as it is science. Now, the mathematical and aesthetical intricacies curve-fitting are beyond the scope of this thesis, but luckily are irrelevant to our philosophical discussion. We shall dwell on the fact, already noted before, that postulation of *any* functional relation between measured quantities as a result of curve-fitting is a form of *ampliative inference*. From our limited body of evidence we infer a hypothesis that represents more than just a summary of the evidence: it augments our knowledge by yielding claims about past data (retrodiction) as well as future data (prediction).

This sort of ampliative inference is a form of idealization, because the world is a messy place. Most of our empirical data is obtained through a process that contains an inevitable margin of error due to measurement inaccuracies, interference and other contingent anomalies we have no control over. What we are looking for in making measurements is, however, the "uncontaminated", true relationship between the quantities measured. We infer the true relationship from our data by interpreting that data, through a *model*. To be able to do anything at all with our data, in particular to predict new data from it, we need a model. In the words of Stephen Norton and Frederick Suppe: 'Neither raw data nor raw sensory experience carry their own interpretations. To be properly interpreted and deployed, data must be modeled. As one researcher put it, "If you ask a scientist for her data, she just gives you a model"'(Norton and Suppe 2002). Any set of actual data, no matter how large, will be finite, and due to the above mentioned factors will never faithfully represent the true relationship between the quantities measured. When fitting a curve onto that data set we will attempt to ignore what we presume to be the effects of measurement error, etc., and approximate what we presume to be the "true curve." Consequently, curve-fitting is a process of finding a function that, although it closely approximates our data concerning the relation between the variables—after all, the data is often our only direct evidence concerning this relation—is not *too* close to the actual data. On this sort of picture, curve-fitting is one of the ways in which we separate the (complex) "noise" from the (simple) "signal", or eliminate disordered appearances by formulating hypotheses concerning the true, and well-ordered, forces in nature.

The straight line over the data in our graph is thus presumed to be the functional relation between x and y, which we *should* have measured, in the sense that it is what we *would* have measured if we had been able to eliminate measurement error and other interference. The straight line says "this is how measurements would be," if conditions were ideal. In many cases, it is the graphic representation of the set of predictions and retrodictions derivable from an ideal law. For instance, our favourite example of an ideal law, the Ideal Gas law, can be rewritten as $P = R\frac{NT}{V}$, where P, the dependent variable, is a linear function of $\frac{NT}{V}$ and constant R. The straight line given by $P = R\frac{NT}{V}$ approximates the data obtained from actual gases.[130] It is an approximation, but also an idealization, in so far as we know beforehand that no actual measurements should ever yield the straight line exactly. If they do, then we know that that must have been a rare instance of several interfering factors (the noise) cancelling each other out so as to, coincidentally, yield the straight line. The utility of finding the straight line through curve-fitting is precisely that it represents just the functional relation between the measured quantities, and ignores further interfering factors. The law it stands for says "if the causal factors a, b, c, … were absent, then a quantity of gas N in a container would behave according to $P = R\frac{NT}{V}$". We do not have to look at the way things are, actually, to know that a, b, c, … are always present.

3.2.2 Idealization *vs.* approximation

Ever since Galileo inaugurated experimental science by attempting to verify his law of Free Fall through calculating its consequences for 'a *perfectly round* ball rolling down a *perfectly smooth* inclined plane', and then measuring the behaviour of real balls rolling down real planes, the difference between the idealized and the actually observed appeared to many as one of *degree*. Galileo did not of course dispose of a perfectly round ball, nor of a perfectly smooth plane, but he knew the importance of minimizing the distance between the idealized and the actual, so he used a hard and smooth bronze ball, and let it roll down a groove that was as straight and smooth as they could be made at the time. His underlying assumption must have been that the smoother and harder the ball and the plane on which it rolls—i.e. the more experimental conditions are "idealized" by eliminating or reducing external causal factors such as friction—the more measurements will approach the values predicted by the law. McMullin describes the "method of Galilean idealization" as follows: 'The move from the complexity of Nature to the specially contrived order of the experiment is a form of idealization. The diversity of causes found in Nature is reduced and made

[130] For more precision on this point, see Footnote n° 131.

manageable. The influence of impediments, i. e. causal factors which affect the process under study in ways not at present of interest, is eliminated or lessened sufficiently that it may be ignored.' (McMullin 1985, at p. 265).

Galileo inaugurated an approach to idealizations and their confirmation based on the idea that idealizations are kinds of *approximation*: many, in Galileo's wake, have considered idealization an act of theorizing that produces approximately true theories or laws such that, were the idealization removed, the resulting law would bring us closer to the truth. On this sort of approach—which with some justice can be called the traditional approach to idealization—it is expected that, generally, experimental results and predictions derived from an idealized law *converge* if experimental conditions are improved, or, alternatively, if the idealized law is relaxed and made more "realistic" (cf. Liu 1999, p. 239).[131]

This idea is a rather intuitive one. If idealization is indeed an act of constructing an approximately true theory or hypothesis about reality that eliminates or reduces interfering factors, then as we 'bring the analysis to the phenomenon' by removing or weakening the idealization, we loose whatever advantages in simplicity of description or conspicuousness it brings with it, but gain a closer description of com-

[131] Cf. also Laymon, R. (1985). "Idealizations and the Testing of Theories by Experimentation" *Observation, Experiment, and Hypothesis in Modern Physical Science*. P. Achinstein and O. Hannaway, Cambridge, Massachusetts, MIT Press, and Laymon "Idealizations". Laymon gives the following example for this sort of convergence: the equation (1) $ml^2\frac{d^2\theta}{dt^2}=-mgl\theta$, establishes an idealized relation between a given pendulum's mass (m), the length of its suspension cord (l), its angular displacement (θ), the downward pull of gravitation (g), and the pendulum's period (t). Laymon points out that we can render this equation progressively more descriptive of the behaviour of real pendulums by peeling off one "layer" of idealization after the other, or by reducing the "degree" of the idealization. Thus, we can first reinstate 'sin θ' for 'θ' ('sin θ' had been dropped in order to facilitate computation, justified by the "negligible" difference, for small θ, between θ and sin θ), thus properly accounting for the oscillating movement of the pendulum: (2) $ml^2\frac{d^2\theta}{dt^2}=-mgl\sin\theta$. Then, we can add a coefficient of resistance to account for friction: (3) $ml^2\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}=-cl\frac{\mathrm{d}\theta}{\mathrm{d}t}-mgl\sin\theta$. $\langle(1), (2), (3)\rangle$ represents a sequence of steadily more realistic, or less idealized, mathematical descriptions of the movement of a simple pendulum. As we move from (1) to (3), our predictions derived from this sequence will converge towards our measurements of real pendulums. Laymon holds that this example of gradual convergence between predictions and observations is typical for the sciences (Laymon "Idealizations", section 3).
The movement of convergence as idealizations are "relaxed" in this way is comparable to what Nowak, L. (1980). *The Structure of Idealization*, Dordrecht, Reidel, and others in the "Poznan School" call the process of 'concretization.' According to Liu "Approximation, Idealization, and Laws of Nature", p. 239, Nowak's account can be understood as a special case of Laymon's. For a succinct discussion of Nowak, cf. Cartwright *Nature's Capacities and their Measurement*, pp. 202ff. For a discussion of how 19th-century scientists and engineers attempted to "correct for" the idealizations involved in (1)—in particular for resistance factor c—and build practically usable instruments for measuring fluctuations in gravity, see Laymon, R. (1989). "Applying Idealized Scientific Theories to Engineering" *Synthese* **81**(3): 353-371, at pp. 355-58.

plicated, messy reality. Conversely, if we ‘bring the phenomenon to the analysis’ by artificially manipulating experimental conditions, we loose closeness to reality, but gain in the degree of fit between predictions derived from the idealization, and actual observations. The ‘ideal’ appears on this sort of picture as the limit of the ‘real’, and it is natural to expect convergence. Ronald Laymon has seized upon this feature to develop a theory according to which an idealization is viable just in case predictions derived from the relevant idealized law/theory converge monotonically on observations (Laymon 1985 and Laymon 1987).[132] *Monotonicity* is the requirement that the degree-of-fit between predicted outcome and actual outcome improve in the same direction each time the idealization is made more realistic, or alternatively each time experimental conditions are modified to better resemble the idealized condition. On this ‘approximation-improving’ view of idealizations, whether we have used the right kind of idealization depends on whether by diminishing the distance between the real and the ideal, we produce just this kind of convergence.[133]

Applying this criterion to Fodor’s claim that our attribution to Don of the disposition to add (D_A) represents a legitimate case of scientific idealization means that there ought to be monotonic convergence towards actual observations as we remove, one by one, the idealizational assumptions implicit in that ascription. Now, suppose the relevant idealized law is

> Agent x is disposed at time *t* to add iff x has the complex property G at time *t*

or $D_{A,t}(x) \equiv G_t(x)$, where we analyze $G_t(x)$ as

> $G_t(x) \equiv CP$ (If C_i, then if $S_t(x)$, then $M_{t'}(x)$).

[132] Interestingly, Laymon “Idealizations” refers to his equation (3), i.e. the more complex phenomenological description of the behaviour of pendulums as they are affected by interference, as “the truth”. Laymon’s stance that idealized laws are strictly speaking false is not uncontentious; one might very well argue that idealizations, qua idealizations, are neither true nor false, because they are not *bona fide* representations of reality in the first place. (cf. for example, Hüttemann, A. (1996). *Idealisierungen und das Ziel der Physik: Eine Untersuchung zum Realismus, Empirismus und Konstruktivismus in der Wissenschaftstheorie*, Berlin; New York, de Gruyter). In our discussion of curve-fitting above, we followed the lead of Forster and Sober “How To Tell...” and called the plotted curve the “true” curve, although it is merely an approximation of the observed data. The question whether the only thing that can be called ‘true’ is the curve, or the data (or neither), is one way to frame the debate between realism and anti-realism in this domain.

[133] For difficulties with this view, see Liu “Approximation, Idealization, and Laws of Nature”, and *infra*.

(X has a property such that, *ceteris paribus*, if X undergoes stimulus "$m + n = ?$" under conditions C_i at time *t*, X responds, at time t′, with the sum of *x* and *y*). In other words, agent x is disposed to add if and only if he could, under ideal conditions C_i, add any two numbers *m* and *n*. Suppose further that 'C_i' denotes a series $\alpha_1 \dots \alpha_n$ of distinct ideal conditions. Monotonic convergence then implies that as we remove α_1 and replace it with a more realistic description of initial conditions, we ought to obtain a prediction from the modified law about the behaviour of real agents that approximates better our actual observations than the unmodified law. Similarly, replacing α_2 should again further diminish the distance between prediction and measurement, and so should replacement of α_3, α_4, etc., until α_n has been removed, and we have obtained a fully empirically adequate phenomenological law about the observable behaviour of agents who are disposed to add. This would be a law whose curve goes through all measured data-points, and correctly predicts new data. Of course, even this law would not necessarily account for unforeseeable kinds of interference, i.e. it would still be CP. In the reverse kind of movement, as experimental conditions are made to more closely resemble C_i, monotonic convergence similarly implies that observations under the improved conditions would have to better approximate the predictions of the ideal law. Thus, Galileo attempted to 'bring the phenomenon to the analysis', in Laymon's phrase, by using a ball and wooden plane that were as hard and smooth as possible.

Now, as pointed out before, in the case of the disposition to add, any improved description of C_i promises to be a non-trivial affair. Fodor believes that part of C_i is 'unbounded memory'; Kripke thinks a larger brain and a longer life are as well (Fodor 1990, p. 95; Kripke 1982, p. 27). This seems reasonable, insofar as even a cognitive system with an enormously large short term memory would still need a non-zero amount of time to actually perform the computation (there is no reason to grant such a thing as instantaneous computation). However, if a brain's volume—or rather that of the frontal cortex, currently thought to be the part of the brain involved in providing short-term memory for various cognitive tasks—and its life expectancy have any causal influence at all on the agent's capacity to add progressively larger numbers, then it is clear that the capacitiy to add *indefinitely* large numbers would require brain sizes and life expectancies that are also indefinitely large. Holding something else would require implausible assumptions. For, suppose that it was the case that the ability to add 10^3-digit numbers required an increase in brain volume of 50% (everything else being equal). It is hard to see why a further increase in the size of the arguments, say to 10^{5-}digits, would then *not* also require a non-infinitesimal further increase in volume, and so on indefinitely. *Ditto* for life expectancy. In other words, there is no reason to assume that requirements on brain size and longevity would asymptotically approach a limit as the size of the computed numbers increases without bounds.

Surely, *indefinitely* large brains with *indefinite* longevity complicate things, if our task is to bring the analysis to the phenomenon, or to bring the phenomenon to the analysis, and look for convergence. We have seen that in the case of the Ideal Gas Law, 'If C_i, then *PV* = *RNT*', ideal condition C_i consisted of at least 2 distinct constituent idealizations, namely perfectly elastic molecules and absolutely impermeable containers, and that these could themselves be decomposed into further idealizing conditions. For elasticity, this was e.g. the condition that the electrical force of a molecule must have a negligible range with respect to its average distance to neighbouring molecules; a further constraint is that molecules must have no internal structure, no parts with internal degrees of movement. The next question, whether the assumption of negligible range for electrical force conceals any further idealizations is probably not physical, but metaphysical—in any case, hard to answer. Similarly, it is quite probable that the requirements of indefinitely large volume and life expectancy conceal several constituent idealizations, or layers of idealization, which we are unable to identify. Bringing the analysis to the phenomenon would require their identification, yet we do not know what could make brains live indefinitely long, nor do we have a grip on the sort of causal factors that interfere with brains being indefinitely large. Even without this sort of knowledge, however, it is plausible that if nature imposes upper bounds on the sizes of atoms as well as on those of planets, stars and galaxies, so brains too are likely to have intrinsic size-limits that are inherent in their architecture and the materials they consist of. They also have extrinsic limits. Any infant's head having had to pass through its mother's pelvic opening, evolution puts an upper bound on the size of brains: any infants with larger heads and brains do not survive birth, and thus cannot pass on their tendency for larger brain size. Going beyond this sort of limit in an idealization would require either the assumption that future human generations develop and widely use extra-uteral methods of child bearing, or that the average size of females also increases indefinitely… We see that the idealizational assumptions implicit in indefinite size are rather substantial indeed. As Kripke put it, we are firmly in the realm of science fiction. *Idem* for indefinite life expectancy.

This is related to an important difficulty for the approximation improving-view of idealizations, mentioned by Laymon himself. He invites the reader to imagine a scenario in which

> ... there are one hundred forces, f_1 through f_{100}, all aligned along a common axis, operating on some particle. And let the first ninety-nine be of equal magnitude but alternating direction, and the remaining force be twice the magnitude of the others. Now imagine that scientist S arrives on the scene, gradually learns of the existence of the forces in the sequence $\langle f_1, f_2, \ldots f_i, f_{i+1}\rangle$, and applies standard Newtonian theory, that the acceleration of the particle will be equal to the product of its mass and the vectorial sum of the forces. … The resultant summations can naturally be described as becoming increasingly more realistic and less idealized. (Laymon 1998, sec. 3)

The problem is, of course, that learning about the existence of force n°1 to 99 will not lead to gradual convergence. Rather, "convergence occurs all at once" when we have finally come up with a fully adequate description of all forces present. The absence of monotonic convergence here is not an indicator of the falsity of the theory, however, nor of the inappropriateness of idealizing away from component forces. Conversely, we can imagine a scenario in which the application to our theory of a long series of gradually less and less idealized input conditions *does* lead to monotonically converging predictions. *Prima facie*, we would be inclined to say that this lends credence to the theory. However, convergence would not constitute evidence of our theory being true, if we were not independently assured that our input conditions (idealizations) are in fact part of the strictly monotonically ordered sequence of correct idealizations for this domain, i.e. that our idealizations are ordered in such a way that the application to them of the *true* theory for the domain will result in strictly monotonically convergent predictions. But how, Laymon asks, are we supposed to know that our idealizations are so ordered without knowing, *beforehand*, that our theory is true? Laymon concludes that theories can only be experimentally tested by assuming that the corresponding idealizations are monotonically ordered, but that idealizations can only be ordered if the corresponding theory is assumed. This, he argues, seems to leave us with no procedure at all to simultaneously test both theory and idealizations (Laymon 1998, Ibid.).

The theory and set of idealizations presently under scrutiny are that most of us are bearers of the disposition to add, and that, *ceteris paribus*, anyone with such a disposition will compute indefinitely large numbers if conditions C_i are optimal. Laymon's argument shows that even if we disposed of a rather long sequence of gradually less and less idealized descriptions $\langle C_1, C_2, C_3, ...\rangle$ of C_i, and if these descriptions yielded steadily convergent predictions when applied to our theory, this convergence would not constitute evidence in favour of the truth of the idealizations, or of the theory. For, given that we do not want to presume true our theory, we would need a strictly monotonically ordered sequence of input descriptions to confirm it. Part of C_i are indefinitely large working memory, *a*, indefinitely large brain, *b*, and indefinitely large life expectancy, *c*. As we have argued, idealizations $\langle a, b, c\rangle$, are situated on a *scale* of idealization that could be called "macroscopic", i.e. under closer scrutiny they are likely to each reveal several underlying layers of further idealized conditions, $\alpha_1 \ldots \alpha_n$, $\beta_1 \ldots \beta_n$, $\gamma_1 \ldots \gamma_n$, which, at present, we know little of. In particular, we do not know how to monotonically order these conditions. If Laymon is correct, then the fact that we cannot construct a full monotonic sequence of the relevant idealizations implies that we cannot experimentally test both the truth of the theory, and the legitimacy of the idealizations. The mere fact that, say, artificial "cognitive" agents such as computers tend to be able to solve larger problems as we increase their working memory (i.e. the mere fact of partial convergence), would not be an indication of either the viability of the idealization, or the truth of the hypothesis that they are indeed adding (instead of, say, quadding).

Unfortunately, the method Laymon tentatively endorses for getting us out of this predicament cannot be applied in our case. Laymon describes a kind of bootstrapping procedure, whereby we first use theory and experimentation to tentatively order the idealizations used in some particular domain of application, and then look at a different domain of inquiry, a domain that '... must be one that is not so different that the previously obtained measures of relative realism will not be expected to hold. If the new theoretical predictions do not retain the ordering of the original domain (assuming appropriate compensation can somehow be made for relevant differences), then suspicion is cast on the theory' (Laymon 1998, sec. 3). In other words, we expect the idealizations, and hence the corresponding model, to generalize to predictions in a different domain. Applied to our case, this means that as we measure the relation of short term memory to the size of successfully computed addition problems, we would tentatively assume that as memory size approaches infinity, the agent tends to be disposed to compute numbers that also approach infinity, and that he would still compute them in accordance with the addition table (we would, in effect, fit a straight line over our data). Looking for confirmation of both this theory and the relevant idealization, we might turn to a different, but not wholly unrelated, domain, namely linguistics, and see whether increased short term memory there also leads to an ability to parse increasingly larger sentences, in which case we might consider both theory and idealization confirmed. But, to flog the horse once more, the problem with this is that we have insufficient background knowledge about interfering factors and anomalies that may or may not enter the fray as brains get inordinately large or old. How do we know if the model, which might work in the narrow domain of memories of a certain limited size, can be extrapolated in this way (for more on this, see Section *infra*)? As we split up our idealization of indefinitely large working memory into its constituent parts, we find ourselves incapable of even tentatively ordering the resulting idealizations with respect to monotonicity. Under these circumstances, and within the framework provided by Laymon, holding that the use of unbounded memory in linguistic theories constitutes justification for the use of that same idealization in psychology, and *vice versa,* would be a instance of the blind leading the blind.

3.2.2 Curve-fitting and Approximation. The Akaike Information Criterion (AIC)

The problems with Laymon's approach bear directly on the issue of curve-fitting. Laymon's notion of 'approximation' is one of *factual* closeness—it is what we measure by the SOS-value in curve-fitting. Chuang Liu points out that this value is not always the approriate measure when it comes to approximation (Liu 1999, pp. 240-44). Suppose a law L, relating quantities F and G, is the true law for a given domain. Then, Liu notes, there will be statements L_1 and L_2 such that L_1 is *closer* than L_2, from the point of view of its SOS-value, to both L and our observations of the relationship between F and G (Liu 1999, p. 241): [134]

[134] Liu does not represent the data in his graph, and goes on to discuss law-likeness without mentioning the problem of curve-fitting. (see *infra*)

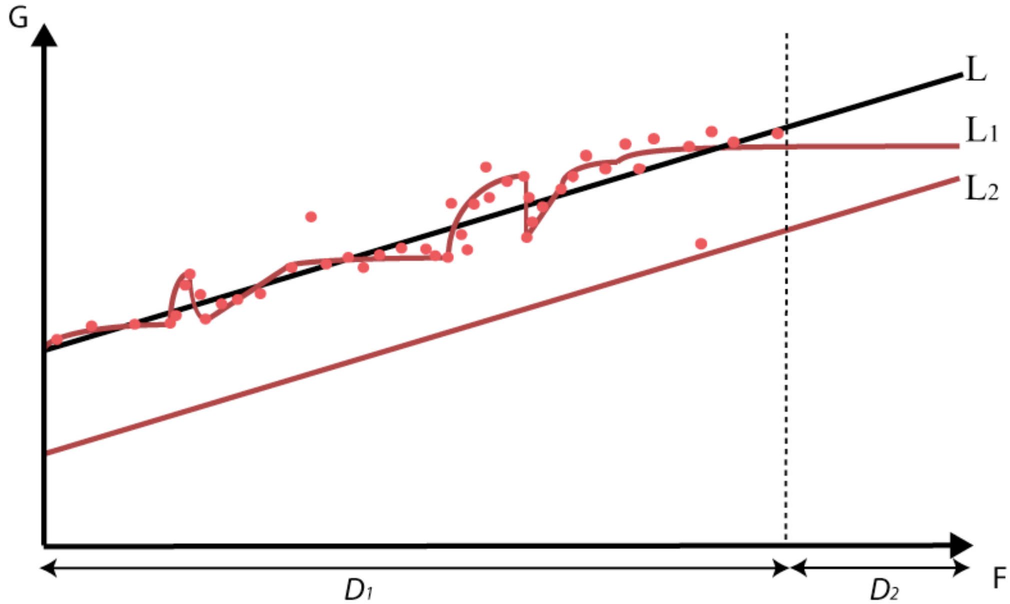

Fig. 12

On Laymon's conception, L_1 would count as the better approximation: it yields predictions that are factually closer to our actual data—it is even closer to the data than the retrodictions of the true law *itself*. Yet there will be no doubt in any experimenter's mind that L_2 is the more *law-like* statement, and that it is, in *that* sense, a better approximation of both L and the data. As Liu puts it, L_1 may be a better approximation *qua* facts, but not *qua* law (Liu 2001b, p. 3).

This is a problem of curve-fitting analogous to those already presented above (L_1, has a lower sum of squares than L itself, and thus "overfits" the data). What is new is the desideratum of law-likeness. Assuming that the correlation of F and G is due to the operation of a law of nature, we want our fitted curve to combine closeness to observational data with law-likeness—where being a law or like a law consists in more than just avoiding over-fitting the data. We expect laws to be of a certain *form*. Therefore, approximating the true law, or being close to it, consists of two dimensions, holds Liu, closeness to the form of the law as well as closeness to any singular observations predicted by the law (Liu 2001b, p. 3). Figure 12 shows that the two criteria do not always yield the same verdict. How do we establish a balance between them? Well, the most striking feature of the form of L_2 when compared to that of L_1, is of course its *simplicity*, and there are well-known curve-fitting methods for establishing a balance between closeness-of-fit and simplicity.

Malcolm Forster and Elliott Sober, for example, advocate the virtues of a statistical theorem, the so-called Akaike Information Criterion (AIC), for its capacity to yield a curve-fitting criterion that allows us to identify, given a particular set of data, the family of curves with the highest estimated accuracy of prediction of future data. The reasoning underlying AIC departs from the observation that the usual requirement of approximating as closely as possible our data—the empiricist's desideratum of staying true to observation)—typically leads us to plot curves that over-fit our data set by needlessly taking into account measurement errors, interferences, and other anomalies. All-too-close curves, such as L_1 in Figure 12, mistakenly take on board all those random factors that make a finite set of data "noisy", and thereby end up making inferior predictions. We therefore need to balance closeness-of-fit, as measured by the method of squares, with the degree of complexity of the relevant family of curves, which is measured by the number of adjustable parameters in its functional expression. A 'familiy' of curves is a set of curves given by a function with one or more such parameters. For example, all straight lines have equations that are of the same form but which have different values assigned to one or more parameters in the equations:

$$f(x)=ax+b$$

The family of straight lines thus has a degree of complexity 2, for it contains adjustable parameters 'a' and 'b'. The family of parabolas, on the other hand, is given by

$$f(x)=a+bx+cx^2$$

and has a degree of complexity 3, as it contains three such parameters. Our discontinuous curve representing L_1 in Figure 12, having been constructed out of (sections of) ten different parabolas, is member of a family of curves with a degree of complexity of at least 30.

Any *particular* curve plotted in the graph is the result of estimating the best value, given the data, for each of these parameters to obtain the one curve with the closest possible fit. Our real curve-fitting problem, on this approach, is not so much the problem of fitting a particular curve onto the data (and hence of defining simplicity for a particular curve), but rather that of finding the right family of curves, or model, for the data among infinitely many alternative models (Forster and Sober 1994, p.). In a follow-up to that paper, goes out of his way to make it clear that AIC does *not* define simplicity for particular curves, and does not even tell us how to fit particular curves.[135] Once the model has been chosen according to AIC, the "best-fitting member" of the model needs to be determined by calculating the effect of varying the value of each parameter on the sum of squares. In our Figure 7 above, a glance at the data suggested to us that the family of straight lines must be the best model for the data; to find the best-fitting member of that family of curves, we would need to estimate the optimal values for each parameter.

Our main problem is hence the initial choice of *model*, or family of curves. Forster and Sober point out that this choice was traditionally conceived as conceptually quite separate from the actual process of fitting a curve. For whereas the result of any given application of the method of squares is determined, once the choice of model has been made, by the nature of our particular data-set—that choice itself may seem somewhat *a priori*. It is easy to think that it is governed by a general preference for simplicity which is either merely pragmatic (e.g. because it makes computation easier), or the upshot of what is ultimately a *sui generis* preferability of simple over complex hypotheses (e.g. because we think that the universe is, or must be, simple). Forster and Sober locate the philosophical significance of AIC precisely in the fact that it renders pragmatic or metaphysical considerations regarding simplicity superfluous by showing how we can conceive of the preference for simplicity as tightly integrated with the process of determining the best-fitting curve. For, the fewer adjustable parameters a model contains, the fewer values need to be estimated (in order to apply the method of squares or some other technique). The fewer values need to be estimated, the smaller the estimation error, and the better predictive accuracy. Our preference for simplicity can hence be justified by *the data itself*, according to Foster and Sober, to the extent that AIC shows that, given the data, a simpler hypothesis will have the tangible effect of tending to lead to better predictions of future data (Forster and Sober 1994, p. 27).[136]

Here is what AIC explicitly says: given a certain data set, we obtain an estimate of a given family's predictive accuracy by adding three terms, (1) the sum of squares of that family's best-fitting member, (2) the number of adjustable parameters

---

[135] Cf. Forster "Model Selection in Science: The Problem of Language Variance". According to Forster, this fact, together with the caveat that AIC's criterion for the simplicity of models is not *literally* the number of adjustible parameters, either, serves to defend AIC from the charge of language-relativity (see *infra*).

[136] AIC is compatible with a realist conception of science, they add. It shows how science "aims at the truth like an archer aims at the bull's eye, even if he has no hope of hitting it" (Forster, M. and E. Sober (1994). "How to Tell when Simpler, More Unified, or Less "Ad Hoc" Theories Will Provide More Accurate Predictions" **45**: 1-35, p. 28).

of that family multiplied by 2 and term representing the distribution of errors, and (3) a constant,

> Estimated (Distance from the truth of family F) = $SOS(L(F))+2k\sigma^2+\text{Constant}$
> (Forster and Sober 1994, p. 9)[137]

The second term is the crucial one, for it is the one that adjusts for "overfitting." The claim is that AIC provides us with a straightforward, quantitative, and reliable method for ruling out spurious candidates such as L_1, which it penalizes for their high degree of complexity. Although defeasible in any one particular instance, the theorem is thought to yield an unbiased estimate in the long run (as we multiply data sets) of the predictive accuracy of any given family of curves. Moreover, according to Forster and Sober, AIC shows us how to justifiy the use of simplicity considerations in the evaluation hypotheses as considerations that are *not* extraneous to the data (such as metaphysical or pragmatic arguments). It thus proves that "the data tell you more than [the Empiricists] may have thought" (Forster and Sober 1994, pp. 27-28).

It is tempting to think that AIC provides the solution to all our problems. If it could be shown that the grue-hypothesis manifests a higher degree of complexity than its rival (as some authors claim it does, cf. Harman 1994, p. 159) in the form of an additional adjustable parameter—namely its reference to time—then we could apply AIC to demonstrate the correctness of what we already believe, i.e. that adoption of the grue-hypothesis is liable to lead to false predictions of the colour of future emeralds. Similarly, construing disposition-ascriptions as acts of curve-fitting, we could explain why, given our data, the ascription to Don of the quadding-disposition involves fitting a more complex curve than the ascription of the addition-disposition. AIC could also help to explain why the hypothesis 'Peter is courageous' is preferable to 'Peter is courageous, except on Monday afternoons and Fridays'. Finally, we could get at least a partial grip on idealization, by showing that many acts of idealization—in so far as they achieve approximation through simplification, i.e. through ignoring additional adjustable parameters—are governed by the very same bias for simpler hypotheses as quantified by AIC. Thus, for example, the Ideal Gas Law could

[137] The distribution of errors, σ, measures the degree to which our actual data tends to fluctuate around the true curve. Clearly, a greater number of "outlying" data points, or anomalous observations, will tend to decrease the curve's average predictive accuracy—i.e. its closeness to future data as measured by the method of squares. Incidentally, Forster and Sober also provide a more general statement of AIC in terms of *likelihood* that is applicable to cases for which no sum of squares is defined, using the fact that low SOS value is correlated with high likelihood (Ibid., p. 10).

be reconstructed as the result of an act of fitting a curve onto an initial set of empirical findings concerning the relationship between the pressure, temperature, and volume of a confined gas.[138] The idealizing assumptions of the law could then be interpreted as an upshot of the desire to keep the hypothesis, PV = nRT, free from additional adjustable parameters. Van der Waals' equation, (P + an2/V2)(V – nb) = nRT, does *not* make these assumptions and introduces "correction" terms 'a' and 'b' to account for intermolecular forces and the fact that molecules have non-zero volume. Whether or not AIC would yield the verdict that despite its increased complexity, '(P + an2/V2)(V – nb) = nRT' is in the long run the better curve-fitting solution, depends on the error distribution in the data we happen to be working with. The more accurate our data, the more likely it becomes that AIC would favour Van der Waals, for as a more precise and "realistic" model it would eventually overcome the penalty for complexity through a better SOS value.[139]

Predictably, it has been quickly objected to Forster and Sober that AIC's criterion of simplicity—the number of adjustable parameters—renders it powerless to deal with a problem such as the new riddle of induction, because the number of adjustable parameters is *conventional*. According to Scott DeVito,[140] Akaike's theorem provides a way to choose between hypotheses that is inherently relative to the way we conceptualize the world, as expressed in our choice of coordinate system. By chosing a different coordinate system, we can arbitrarily vary the number of parameters. DeVito deploys, in substance, Goodman's argument of the language-relativity of simplicity as expounded in Section 1.2.2, and as already put forward by Priest 1976. Forster counters that the 'grue'-hypothesis does not, properly speaking, contain any adjustable parameters at all—rather, it contains one *adjusted* parameter. After all, the term '*t'* in the hypothesis is not intended to *vary* across all possible values for *t*, but

[138] In fact, historically the situation was not as straightforward: the more general Ideal Gas Law was derived from "special cases" (1) Boyle's experimentally measured law V′/V = P/P′ (the volume of a gas at a constant temperature is inversely proportional to its pressure), (2) the equally experimentally measured Charles' Law, V ′ = V(1+θ/273 K) (the volume of a gas under constant pressure increases or decreases with temperature), and, Avogadro's hypothesis, V = *aN* (the volume of a gas is directly proportional to the number of molecules of the gas) (Orear *Physik*). We can account for this by assuming that the initial curve-fitting exercise concerned the two experimental laws—similar idealizing assumptions are involved in these.

[139] It is unlikely, however, that AIC would have been of any help whatsoever in *discovering* van-der-Waal's equation, which depends on us already possessing a kinetic theory of gases. Thus, the importance of background knowledge of underlying causal processes at least equals that of statistical methods when it comes to assessing empirical data. (I am indebted to Donald Gillies for this point).

[140] DeVito, S. (1997). "A Gruesome Problem for the Curve-Fitting Solution" *British Journal for the Philosophy of Science*: 48(3) 391-396, p. 392.

rather is thought to be fixed at some remote point in the future. Consequently, we are always dealing with *one* specific 'grue'-hypothesis and *one* corresponding curve, not a family of hypotheses and a family of curves (a model) (Forster 1999, p. 93). Moreover, Forster claims that even if we changed Goodman's original definition by replacing *t* with a genuine adjustable parameter, θ, so that we would always be dealing with a family of infinitely many alternative 'grue'-hypotheses,

> $\forall x$ [x is grue iff (x is green & observed before θ) $\cup$ (x is blue & $\neg$ observed before θ)]

(where θ varies), then this would still not make AIC language-relative, or conventional . However, it would make AIC *inapplicable* to the grue-case (Forster 1999, p. 96).

AIC's *rationale,* according to Forster, for holding the number of adjustable parameters low is the fact that when picking out the best-fitting member of a family of curves, the parameters' values need to be estimated on the basis of the available data. The term '$2k\sigma^2$' attempts to quantify the total sampling error in these estimates, which will tend to increase as the number of parameters increases. However, Forster points out, given the alternative definition of grue, it is in fact both useless and impossible to estimate the parameter 'θ'. It is useless, because no matter which value θ takes on, it will make no difference to degree of fit of the hypothesis with existing data (as long as that value lies in the future, which, by assumption, it does). Hence no particular value for θ will allow us to individuate the best-fitting curve in the family of grue-hypotheses. They all fit equally well. Indeed, another look at our curve-fitting problem,

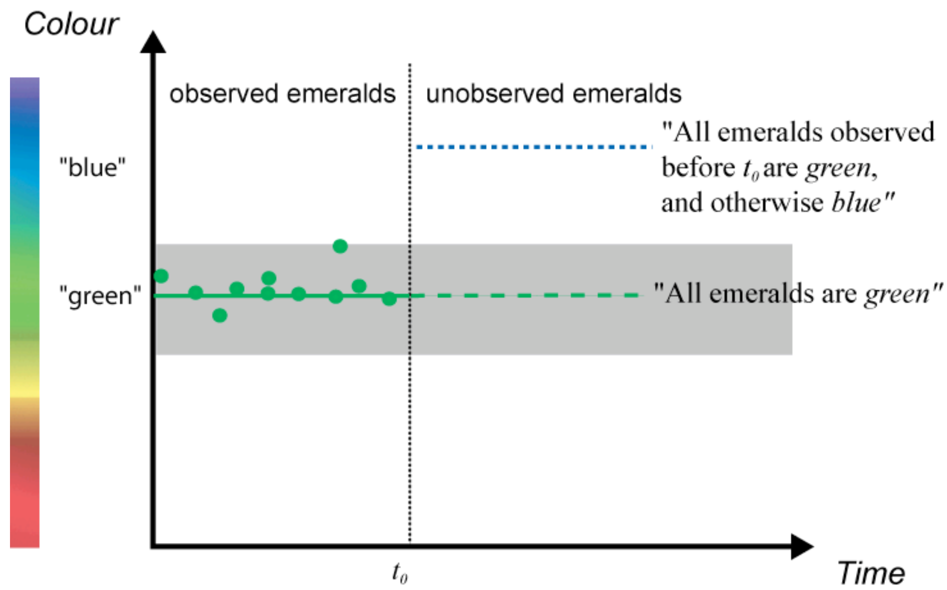

Fig. 13

reminds us that the rival hypotheses "All emeralds are *green*", "All emeralds are *grue*" and all its cognates, if taken to express a function from time to the colour of emeralds, achieve an equal fit with our *actual* data-set, which is comprised of past observations of emeralds. Forster points out that the choice between the green-hypothesis and any member of the family of grue-hypotheses as reconstrued here amounts to taking a view concerning the value of θ: the green-hypothesis puts θ at infinity, observed emeralds will *always* be green, whereas grue-like ones put it at lower values. The *crux* is that estimating the value of θ on the basis of today's evidence is simply *impossible*:

> You cannot estimate θ because you cannot sample from the future. Akaike's theorem has to do with predicting new data from information on old data, where both come from the same distribution. Past observations may tell us where θ is not, but predictions require us to say where θ is. The grue problem (set up as problem in which one estimates θ) is like sampling from one urn and using the information to make predictions about a different urn. Without assumptions about how the two urns are related, this is an impossible task. (Forster 1999, p. 96).

The situation before us is therefore as follows: if grue is defined à la Goodman, AIC is incapable to distinguish between the green- and the grue-hypothesis, incorrectly assigning the same predictive accuracy to both, because it will consider both to be of the same degree of complexity, and will go after degree-of-fit only, which is identical. If grue is defined using an adjustible parameter, so that the grue-hypothesis designates a family of infinitely many hypotheses of the same form (a model), then AIC is inapplicable, because to chose the best-fitting member of that family it requires that the value of parameters be estimable on the basis of available data, which is not the case: there are infinitely many potential values of θ that have *no impact at all* on degree-of-fit, making θ unestimable (Forster 1999, p. 96). The point is precisely that in preferring the green-hypothesis, we do not make any sort of estimate at all, we prefer it on grounds not within the scope of AIC. Of course, it would have been a theoretical achievement of monumental proportions if it had been any other way! We are, after all, back to the problem of induction as we know it, i.e. to the problem of projecting from samples to populations *without* being allowed to make crucial assumptions about how the samples are related to the populations—for example, without assuming that the samples are "representative" of the populations, because nature is uniform.

Forster's arguments for the claim that this sort of problem does *not* make AIC language-relative in the way DeVito suggested are interesting, but do not directly concern us here.[141] The important point for our discussion is his acknowledgement that because AIC essentially quantifies estimation error, *it captures only one kind of simplicity*, one that is relevant to the curve-fitting problem conceived as a problem of

[141] Forster points out that AIC specifies a 'discrepancy function' which estimates, for a given family of curves, the distance between the member of that family that best fits the observed data, and the true curve. The term 'k' in that function does not *literally* speaking denote the number of adjustable parameters of the relevant family of curves, because such a number would indeed be relative to how the family is described. Rather, he holds, 'k' denotes the number of parameters that *contribute to the discrepancy function in a specific manner* (Forster, M. R. (1999). "Model Selection in Science: The Problem of Language Variance"Ibid. **50**(1): 83-102, pp. 99-100.) The idea here is that this *contribution* remains invariant even under those redescriptions that modify the number of parameters. Just as, say, the contribution of 'a' to the function 'y = ax+b' must be conserved by '(c+d)' under the transformation 'a = (c+d)', if 'y = (c+d)x+b' is to continue to describe *the same* function. Thus, functions are correlated one-to-many with descriptions, but if a redescription of a function that has been obtained through the transformation of some parameter is to be a *re*description, it must refer to the same function. This implies that our worries about redescription in natural language do not carry over to the language of mathematics. Language-invariance is "built in" at the very foundation of the mathematical theory of functions, comments Forster, the only problem being 'explaining this to a non-mathematical reader' (Ibid.). One is entitled to wonder, however, whether his cheerful endorsement of a clear distinction between the mathematical formalism and the entities it denotes is not tantamount to simply presuming true one of the available philosophies of mathematics, namely Platonism.

model selection (Forster 1999, p. 94). Evidently, this simplicity is not the kind that distinguishes the green-hypothesis from the grue-hypothesis, for no sort of estimation is at work in our fixing the value of θ at infinity: 'The bottom line,' says Forster, 'is that the Green and Grue hypotheses both fit the current observations equally well. None of the usual model-selction criteria provide any reason to favour the Green hypothesis over the Grue hypothesis, either by differences of simplicity, or by differences in fit, and Akaike's criterion is no exception' (Forster 1999, p. 94). Thus, although AIC does help us in determining why Liu's law-candidate L_2 is a hypothesis that is vastly preferable over L_1, it does *not* help us in saying why L_3 is also preferable to L_4:

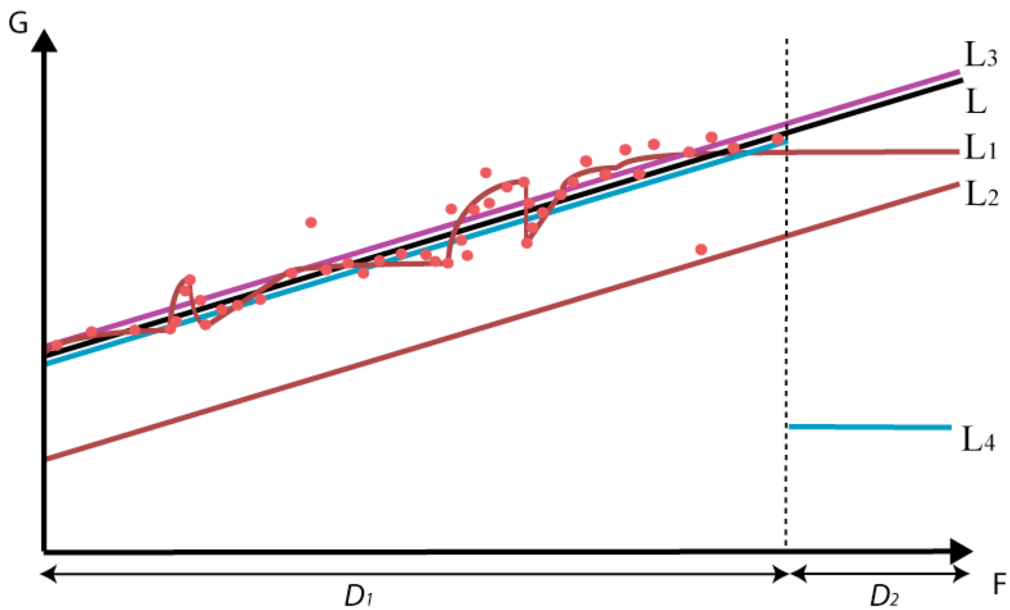

Fig. 14

Within domain D_1, from which our data are drawn, L_4 and L_3 match the true curve equally closely—however, under the assumption that L is the true curve, L_4 performs very poorly compared to L_3 when we use it to "extrapolate" beyond D_1 to make predictions about domain D_2. Of course, L_4 is our grue-like hypothesis here, and we *expect* it to perform poorly on the day it predicts emeralds to be blue.

It would be handy indeed to have a test that predicts such poor performance using *only* empirical data from D_1, sparing us the thankless task to sample from the future by waiting until it arrives. Unfortunately, Forster notes, statistical model-selection criteria such as AIC are not actually concerned with this problem, they merely show us how to find the best mathematical approximation to our actual data by looking at how well it accommodates arbitrarily many additional data-sets *drawn from the same domain*. In other words, AIC provides constraints on what I have been referring to as ‘retrodiction’ and what Forster calls ‘accommodation’, not prediction. Prediction, however, is what would make AIC truly exciting, and it is what a solution of the grue-problem would require.[142]

Forster acknowledges that what we are *really* looking for, in a curve-fitting solution, is the kind of (ampliative) inference that increases our knowledge by licensing predictions, not only retrodictions. In Forster 2000 he reports on the results of fascinating attempts by himself and others to quantify and compare the *degree* to which a whole panoply of different model-selection methods—among which both AIC and the Maximum Likelihood method (the rule that always favours the curves with the smallest SOS-values)—get cases such as L_4 *wrong*. Forster first picks an arbitrary test function, ‘ $y = \frac{1}{2} + \frac{1}{2}\tanh(x-2)$ ’ which generates what will be considered the “true curve” in his test case, and takes a look at how closely various families of functions of a certain form (“models”) manage to approximate that curve for x-values 0 to 3.5. He notes that although certain exponential functions peform even better, given a sufficiently large data set the family of 4-degree polynomials—of wich the family of straight lines and parabolas are sub-families—will achieve quite a good fit for $\{0 \leq x \geq 3.5\}$ (Forster 2000, p. 227; see Figure 15). Interestingly, this changes radically if one widens the domain to $\{0 \leq x \geq 5\}$, where that same model fails conspicously:[143]

[142] This sort of problem is directly related to the wider debate between realists and anti-realists in the philosophy of science. As Laudan, L. (1981). “A Confutation of Convergent Realism” *Philosophy of Science* **48**: 19-49, p. 20, points out, many forms of ‘convergent epistemological realism’ make the double assumption that (a) “mature” scientific theories are typically approximately true, more recent theories being typically closer to the truth than older ones (which is what makes this realism ‘convergent’), and (b) if a theory is approximately true, then it follows that it will be a relatively successful predictor and explainer of observable phenomena. Laudan registers the complaint that ‘promises to the contrary notwithstanding, none of the proponents of realism has yet articulated a coherent account of approximate truth which entails that approximately true theories will, across the range where we can test them, be successful predictors’ (Laudan “A Confutation of Convergent Realism”, p. 32).

[143] The curves in Figure 15 have not been computed and merely approximate those given by Forster.

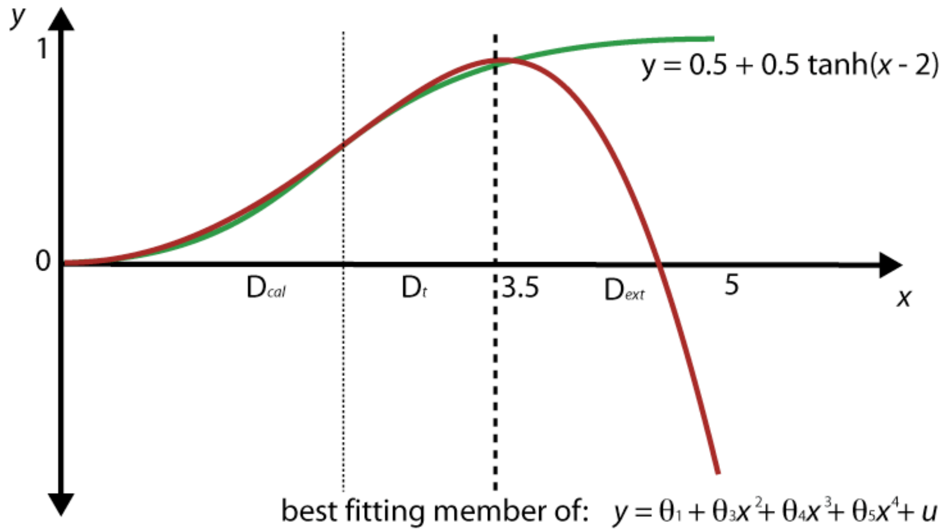

Fig. 15

Of course, if we were provided with data from $\{0 \leq x \geq 5\}$, we could adjust the values of the parameters in such a way that the best member of the polynomial model again achieves relatively good overall fit—in other words, although it fails dismally at prediction, the model is entirely capable of *accommodating* the data, i.e. of retrodiction.

Suppose '$y = \theta_1 + \theta_2 x + \theta_3 x^{2+} \theta_4 x^3 + \theta_5 x^4 + u$' were our grue-like hypothesis, and $\{0 \leq x \geq 3{,}5\}$ the domain from which our past and present observations were drawn. Are there any *empirical* clues, somehow buried in our data from that domain, about which functions will or will not successfully extrapolate beyond it? Forster notes that *all* model-selection methods in his survey—AIC, Maximum Likelihood, as well as, interestingly, 'BIC', a method derived from Bayesian probability theory according to which models should be evaluated by their posterior probabilities—fall short in this regard, insofar as they do not predict the failure of the polynomial model (Forster 2000, p. 228).[144] There is, however, light at the end of the tunnel, Forster suggests, thanks to the 'generalization criterion methodology' (Busemeyer and Wang 2000). The idea here is to judiciously divide the domain of our actual data into a "calibration" domain D_c and a "test" domain D_t, and to fit a curve onto the data in D_c. By comparing the measure to which a close fit of the curve in D_c is an indicator of a close fit in D_t we "test" the curve's aptitude at extrapolation. Success at the test extrapolation is then taken to be an indicator of a likely success at the wider extrapola-

[144] For the specific problems of Bayesian probability theory with the grue-paradox, see Sober, E. (1994). "No Model, No Inference: A Bayesian Primer on the Grue Problem" *Grue! The New Riddle of Induction*. D. Stalker, Chicago, Open Court.

tion to D_{exp} (Forster 2000, p. 229; see Figure 15). Forster comments that

> Such methods rest on the simple inductive argument that if extrapolation has been successful in the past, then extrapolation will be successful in the future. Of course, there is no guarantee that nature will cooperate in this regard; but for that matter, there are no guarantees for the success of predictions of any kind. The issue is not whether such an argument is fallible, but whether there are situations in which past extrapolation is a useful indicator of future extrapolation…’ (Forster 2000, p. 226)

The real life usefulness of the method turns out to be rather mixed, by Forster’s own admission. Although it predicts better than the other methods how the various subfamilies of the polynomial model fare when confronted with the task of extrapolation to $\{0 \leq x \geq 5\}$—it even correctly yields the result that the family of straight lines extrapolates best of all—the generalization test, too, fails to predict how badly the best-fitting member of ‘$y = \theta_1+\theta_2x+\theta_3x^{2+}\theta_4x^3+\theta_5x^4+u$’ generalizes (Forster 2000, p. 229). Forster concludes that the criterion appears to pick up on *some* useful empirical information that is not exploited by the other selection criteria, and that in some situations at least (he does not claim that his results can be generalized), such information might supplement the latter (Forster 2000, p. 229).

Light at the end of the tunnel there may be for *some* curve-fitting problems, but not for those that are *grue*-like. No matter how skillfully we divide up our observations into calibrating and test domain,

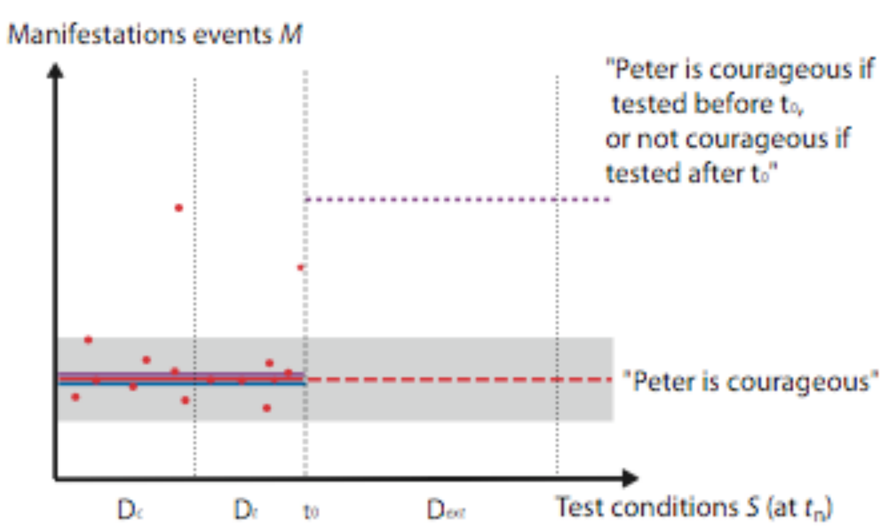

Fig. 16

both the "straight" as well as the *grue*-like hypothesis accommodate the data equally well, and they both extrapolate equally well from D_c to D_t. Not only do we not have any guarantee that if extrapolation has been successful in the past, then extrapolation will be successful in the future (guarantees are not to be sought in evaluating inductive inferences), *grue*-like cases remain recalcitrant precisely because *no amount* of past successful extrapolation will be a useful, or reliable, indicator of future successful extrapolation. Clearly, successful prediction in these cases would require the method to tell us what law-likeness (or, at least, non-*grue*-likeness) consists in, and this is perhaps too much to ask from any statistical method. No matter how good their SOS-values, we do not need statistical theorems to immediately suspect that *grue*-like hypotheses are not particularly law-like, and that barring relevant background knowledge to the contrary, empirical verification of our disjunctive hypothesis about Peter, or of L_4 in Figure 14, would be low on any experimenter's list of priorities.

3.3 "Carving Nature at its Joints"—Yes, but Which Ones?

3.3.1 Disentangling *vs.* Limiting Case Laws

Liu suggests that we need to take a more nuanced approach to idealization. While approximation, approximation improvement, and convergence are surely part of it, they cannot be the whole story. Indeed, there can be no single conspicuous characterization of idealization, Liu claims—no one logical form of all acts of idealization—for idealizations as they are practiced in science are a rather colourful bunch, made for multiple purposes and applications, to each of which corresponds its own type of *reasoning* (cf., Liu 2001b, p. 4; Liu 2001a, p. 5). This is ultimately why the traditional monolithic approach to idealization must fail:

> The reason for the failure of the traditional theories is that there are different types of idealizations, of which approximation production is only one type—e.g. those idealizations one depends on in finding generalizations in a collection of experimental data. But this is not the same kind of reasoning as the

> one which assumes that the universe contains only the sun and the earth or is without electromagnetic field. Nor are both … the same as the construction of lattice models for bulk matter and for quantum fields. From the perspective of approximation production, we simply do not see anything in common in these different acts of idealizations. We must therefore begin afresh. (Liu 2001a, p. 5).

The "Galilean" conception of idealization as approximation must be jettisoned, which also means rejecting Galileo's modern epigones, e.g. Laymon's 'approximation improving'-approach.

Although he heralds the uses of idealization as multiple, Liu does go on to single out one that is special in his eyes: some idealizations are undertaken with the specific aim of *carving nature at its joints* (Liu 2001b, p. 4). In other words, there is a type of idealization whose mode of reasoning is *ontological* and qualitative, rather than quantitative—the separation in the mind of nature into its true component parts and processes. Interestingly, Liu seems to think that—presumably because of the problems with approximation we have seen above—there is little to expect from further refinement of our statistical tools for data analysis, as far as a proper understanding of idealization is concerned. This is not self-evident. In fact, it seems false to suggest, as Liu does, that "carving nature at its joints" on the one hand, and analyzing and generalizing upon our collected data on the other, are two entirely separate and possibly antagonistic ends, subtended by two modes of reasoning that exclude one another. As we have seen, the problems inherent in data analysis already force upon us the conclusion that idealization is probably more than mere approximation. It is precisely its incapacity to provide workable definitions of lawlikeness and of (the right kind of) simplicity that leads to the failure of AIC when it comes to eliminating grue-like hypotheses. Moreover, it is precisely our pre-theoretical intuition that describing nature in *grue*-like terms would not carve it at its "joints" (would not pick out a natural property), that leads us to reject the predicate in the first place—and it is this intuition about the *locus* of the joints of nature that AIC and the other statistical methods fail to account for. It is reasonable to conclude that the task of carving nature at its joints is already *contained* in the task of choosing the right kind of generalization when analyzing one's data. *Idem* for the choice of the right kind of CP-law (see our footnote n° ..., *supra*)

To Liu's mind, taking idealization as a process by which we discover the true component parts and processes of nature strongly motivates an ontology that includes dispositions as parts of the furniture of the world. For, he claims, assuming that dispositions are the true *relata* of (some of) our physical laws would safeguard the literal truth of most of our laws (Liu 2001b, p. 10). It would also have the added benefit of helping to explain how the appearance that "the laws of physics lie" could have come about in the first place. This is how it is supposed to work: Liu takes the law 'whenever an object M of a mass, *m*, is acted upon by a force, f =*m*g, it moves with an acceleration, g' and points out that generalizations such as these are usually thought to obtain only under the idealization that the relevant object is not being affected by any external force. Given that, strictly, this sort of condition obtains only in a universe devoid of other bodies, the law appears literally false. Liu argues that notwithstanding the fact that the ideal conditions never obtain, we can consider it true if we interpret the law as correlating the actions of forces with *dispositions* to motion. For then we can say, truly, that even in a situation in which M is acted upon by force f *and* an additional force e, it will still be the case that M, being acted upon by f, has at least the disposition to move with acceleration g in the direction of f (Liu 2001b, p. 10). The latter disposition of M to move will simply superimpose with another, caused by force e, to move with a different acceleration in a different direction. The end result—M describing a definite trajectory through spacetime—is the *categorical* display of two dispositions combined. Idealization, in this case, is precisely the act of disentangling the co-instantiations and superimpositions, i.e. the carving of nature at its joints.[145]

The view of idealization as the isolation of entangled dispositional properties sits perfectly with Fodor's account of psychological laws. Recall that Fodor had no qualms with law-like statements quantifying over dispositional states, as in: 'If we

[145] Liu's view of the nature and *rationale* of idealization obviously owes a great deal to Cartwright's work on capacities (cf. Cartwright *Nature's Capacities and their Measurement*, Cartwright, N. (1999). *The Dappled World. A Study of the Boundaries of Science*, Cambridge, Cambridge University Press), a debt he acknowledges. Both authors share an outlook according to which idealization is our reaction to the complexity and irregularity of natural phenomena, and both believe that it consists in "carving nature at its joints" by discovering its true component parts, its capacities or dispositions, and their possibilities of recombination. However, they disagree on whether these latter possibilities of combination are themselves representable in law-statements. In other words, what separates them is a question of the priority of laws vs. dispositions, and the nature of laws. Cartwright has a notion of laws in the spirit of a 'liberalized post-positivist empiricism', as statements describing regularities among properties 'antecedently regarded as OK' (i.e. in most cases, occurrent properties). These regularities that are the product of what she now calls 'nomological machines', or the stable arrangement and repeated combined activity of a set of capacities (cf. Cartwright *The Dappled World*, p. 49). As a consequence, all laws are true merely *ceteris paribus*, namely conditional upon the relevant nomological machine continuing to successfully operate in a stable environment (Ibid.). Liu, on the other hand, thinks that there is no reason why there could not be a strict and literally true law statement correlating modal properties such as capacities—in other words, he has a "non-humean" conception of laws according to which laws are more than mere statements of regularity between observable events, and can contain reference to dispositions or capacities. As we shall see, this interesting and deep question is irrelevant from our epistemological point of view—namely its bearing on scepticism.

did have unbounded memory, then, ceteris paribus, we would be able to compute the value of m+n for arbitrary m and n' (Fodor 1990, p. 95).[146] One might think that focusing on the pure dispositional state of being able to add, without consideration of other capacities and dispositions that enter the fray when that disposition manifests, is a paradigmatic act of isolating properties that are always superimposed in nature (on a par with, say, decomposing a total force into its component vector forces), and that therefore, the relevant idealization is an unproblematic one. Moreover, construing Fodor's putative psychological meaning-law about Don as a law that carves nature at its joints would make it immune to the cricitcism we have levelled against it above, based on Laymon's approximation-improving model of idealization. We pointed out that given the idealization of infinite memory, we will have serious difficulties establishing, as required by the theory, monotonic convergence between predictions derived from the law and empirical measurements. The brain is not likely to behave like a computer that has been fitted with additional RAM, by dutifully computing larger numbers in proportion to the increase in its short term memory (as in Martin and Heil's simplistic analogy[147]—or, rather, what precisely will happen is anyone's guess, with the brain going insane being one likely outcome (Kripke, 1982, p. 27).

By classifying Fodor's law as one intended to carve nature at its joints we are in a position to explain the possible lack of convergence, it seems. For what is attributing to normal cognitive agents the capacity to add if not an act of *isolating*, in thought, that particular disposition from its permanent and necessary superimposition with all other dispositions that normally inhibit its manifestation, and have a permanent impact on the agent's performance? The fact that we mean addition and are disposed to add, but will be eternally unable to fully manifest this disposition in a way that would distinguish it from the disposition to quadd, is thus to be explained on the same model involving a composition of different capacities that we use to explain how the Earth can be eternally attracted by the Sun, without ever falling into it.

However, Liu 2001b has a useful classification up his sleeve of idealizations into two main subspecies, which suggests that the ascription of the adding disposition to humans is *not* in fact an act of carving nature at its joints. Liu distinguishes between (1) causal/dispositional laws that are the results of idealization *qua* the decomposition of nature into its constituent parts and causal processes, and (2) so-called

[146] Strictly speaking, this is of course at best a low-level phenomenological law, or the consequence of one, but we can presume that on Fodor's picture it will be derivable from other, more general, psychological laws that quantify over dispositions, too.
[147] Martin and Heil "Rules and Powers", p. 300—see our quote p. 91, *supra*.

"limiting-case" laws, which, according to Liu, result either from (a) an idealization on the *structures* of the systems on which they operate, or from (b) an idealization on the domains of the law's application (Liu 2001b, p. 14ff). It is tempting to think—and Fodor himself certainly seems to think—that Fodor's psychological law candidate is of type (1), or at least a consequence of a law of type (1)—yet it is of type (2). This becomes obvious if we consider the kinds of law that Liu cites as paradigmatic for his second "limiting-law" category: laws that idealize on the structure of a physical system, (2a), are the Ideal Gas law, mechanical laws of rigid bodies, and hydrodynamic laws of continuous fluids; laws resulting from idealizations on the domain of application, on the other hand, (2b), are all those non-quantum or non-relativistic laws, which can be seen as *limiting cases* of quantum and/or relativistic laws (Liu 2001b, p. 14). What makes both (2a) and (2b) "limiting-case"-laws, according to Liu, is the fact that they are limiting cases of *other* laws: the Ideal Gas law, for example, can be considered a limiting case of van-der-Waal's equation, where we assume that all additional parameters in the equation approach zero; similarly, says Liu, non-relativistic Newtonian laws are limiting cases of the corresponding relativistic laws if we assume the speed and the energy of the physical systems in question close to zero (Ibid.).[148]

One may or may not agree with the appropriateness of Liu's classification for each individual law. Important for our purposes is the sort of fundamental conceptual difference he makes out between (1)-type and (2)-type idealized laws: laws that carve nature at its joints simply separate or disentangle actually existing processes or systems in nature without assuming any *modification* in these processes or systems themselves, whereas limiting-case laws set the value of a specific property of the physical system being modelled to an extreme limit (usually either zero or infinity). This is a clearly intelligible and plausible kind of difference between kinds of idealization, one that is exemplified by at least some of our idealized laws. Equally clearly, Fodor's idealizations such as unbounded memory, unlimited life span, etc., must lead to limiting-case laws in Liu's sense, for stipulating unbounded meory is selecting a property of the relevant

[148] Liu comments: 'In practice, when the gas is diluted and at a high temperature, Boyle's law—just like Newton's law of motion in contrast to the law of motion in special relativity—is a good approximation. In both cases, the van der Waals equation and the special relativistic law of motion are 'truer', rather than true, law statements. In general, it is plausible to understand the idealizations taken on the structures for such laws as taking 'limiting' values of the relevant (or salient) structural properties of the model systems in question. And so, we can also understand the idealized structures in a model as the 'limiting' regime in which a simplified but still approximate version of a true or truer law statement holds.' (Liu "Laws and Models", p. 14). The laws that our idealized limiting-case laws are an approximation of are 'truer', rather than true *simpliciter*, presumably because even van-der-Waal's equation or relativistic laws still literally "lie" in Carwright's sense.

physical system (the cognitive mechanisms of the mind involved in adding), and assigning it an extreme value. Liu suggests that laws of this kind are always approximations of other laws that actually govern the behaviour of the system. Well, no-one has yet found the equivalent of van-der-Waal's equation for psychologal meaning laws—i.e. a more complex, phenomenologically adequate description of psychological agents involved in intentional acts of meaning something by a symbol (or indeed of an agent performing an arithmetical computation)—that successfully identifies and quantifies the relevant interfering factors. But this is of course precisely why we have the idealized law in the first place, such laws are sometimes just very hard to find. Perhaps in psychology we are at the same stage of development as our Kinetic Theory of gases was before the discovery of Van der Waal's equation.

The fact that the sort of law-candidates we are interested in are "limiting-case" laws in Liu's terminology is unfortunate for a realist like Fodor. For it is precisely their being limiting-case type laws that is responsible for the literal *falsity* of most laws in most branches of science (Liu 2001b, p. 15). Only idealized laws of the "disentangling" type are *literally true*, because they relate dispositional properties that are fully instantiated even in actual, non-ideal, circumstances. (It's just that, given nature's messiness, these laws are usually co-instantiated with other laws governing other dispositions, so that they can never show themselves in their full categorical display). This sort of picture is precisely what Fodor was aiming at, and what other realists have in mind as well. Interestingly, the philosopher of science Liu, *qua* realist, presents us a picture of the interaction and mutual inhibition of dispositions and the laws governing them, that is identical in its main lines to the account of metaphysical realists Martin and Heil (see Section 2.2 *supra*). Recall that the master argument of Martin and Heil 1997 was based on the metaphysical distinction between the disposition itself and the set of its (actual/possible) manifestations. The purported consequence was that even in cases where its manifestations are permanently inhibited, the ontological reality, and explanatory relevance, of the relevant dispositional state remains untouched.

The congruence between the sort of view Liu is advertising and Martin and Heil's also shows up in the fact that for Liu, laws of the "disentangling" type bear the title of 'idealized' law only because 'to display them in their real form, one usually needs to construct idealized models.'[149] In other words, disentangling laws are not

[149] *Contra* Laymon, Liu thus maintains that the real connection between idealization and approximation lies not in idealization being a form of approximation, but in the fact that in order to understand the operation of idealized laws (of the disentangling kind), we need to isolate them, and to do this we

true merely in their model, but true *simpliciter*—the model is needed only for heuristic purposes. This sort of "idealized" law is literally true in *this* world and describes its deeper reality. Martin and Heil would be happy to underwrite this view of (some) scientific laws. They, too, would presumably balk at the suggestion that liberating, in thought, a disposition from its state of perpetual superposition with other dispositions and considering the sort of manifestation behaviour it would display in that uninhibited state (or in interaction with just a few select other dispositions) is an act of idealization, if 'idealization' is taken in the almost pejorative sense of a representation that is strictly *false* or only approximate. For, on the contrary, given that for Martin and Heil, dispositions are the ultimate and irreducible constituents of what there is, dissecting a dispositional array and isolating individual dispositions in order to display their specific and permanent contribution to the chaos of manifestations, is to dig down to what is ultimately real, and hence to state literally *true* laws in the deepest sense.

So our psychological law about Don seems to have fallen into the wrong category. Contrary to "disentangling" laws, laws of the limiting-type are *false* in the real world. Moreover, although Liu thinks that the degree of approximation of disentangling laws to the true laws can be determined with precise measure by the theory itself within which the law appears (Liu 2001b, p. 15)—and although this may be true of some such laws, including the Ideal Gas law—it seems certainly overly optimistic for our law about Don, as I have argued above. We may be able to calculate precisely the degree of approximation involved in ascribing to molecules of a certain gas the disposition of absolute elasticity—for instance by obtaining measurements of the loss of momentum incurred by a single molecule after a collision with another molecule, leading to the conclusion that molecules of this gas are, say, exactly 99,5% elastic—but it is impossible to do the same for infinite longevity, memory, etc., for the reasons discussed (see Sec. 3.2.2). What we seem to need is an epistemological theory that determines a non-quantitative measure of approximation in this case, and this is what we haven't got. To elaborate on the parallel with metaphysical realists Martin and Heil, another thing we aren't provided with by the (scientific) realist, just as by the

construct idealized models, which we then *approximate* in experiments (Ibid., p. 17). The idea of laws of nature being strictly true, and of the *raison d'être* of idealization and approximation as the calculation of their consequences for concrete physical systems—i.e. the idea that the *locus* of idealization is in the application of laws, not in the laws themselves—is shared by Earman, Roberts, et al. "*Ceteris Paribus* Lost". They put it to work in a response to the problem of ceteris-paribus clauses: ceteris-paribus laws don't exist, the authors claim, what exists are merely ceteris-paribus statements that we use to characterise our empirical knowledge of the hurlyburly of interacting and superimposing dispositions, forces, capacities, etc.

metaphysical realist, is an epistemology to strengthen our back against the Sceptic.

Interestingly, Liu implicitly acknowledges this, and describes in detail exactly the sort of sceptical conundrum we have been wielding in this thesis against all realists about dispositions. Here is Liu's variant of the sort of case we have been worrying about:

> There may indeed be a more dangerous challenge in this connection that I am probably not able to completely meet. If being sentimental is a disposition of someone, say, Bill, who has not displayed and will never display any emotions which show that he is such a man, then why could not being 'lawyerly' also be a disposition of Bill, where Bill is not, has never been, and will never be, a lawyer; and yet he could have been one had the right set of circumstances obtained at some time of his life. In other words, being lawyerly is a modal property of Bill. The question is this: does it make sense to say that Bill, who is one of the most matter-of-fact man in the world and a banker all his life, actually possesses the disposition of being sentimental and lawyerly, just because (let us suppose) given the right conditions he could have become sentimental and a lawyer? (Liu 2001b, p. 19)

I don't know whether we should rejoice at the spectacle of philosophy constantly hitting upon the same old difficulties, albeit from new angles—or be dismayed by it. In any case, it is clear that Liu's problem case of Bill's lifelong, but unexpressed, lawyerliness, is identical to the lifelong, but unexpressed, non-water solubility of Carnap's match (see Sec. 2.1.1). We know Carnap's solution to *that* problem: if the match was made of wood, and if we have laws correlating being made of wood with not being water soluble—which we have thanks to the fact that other entities made of wood have been subjected to the test condition for water solubility—then we know that the match was not water soluble. We also know what got Carnap into trouble, namely the question of ascribing unmanifesting dispositions to objects for which we have no laws that could link them to manifesting dispositions.

Let's look at Liu's diagnosis of the difficulty with Bill's lawyerliness:

> This to me seems more of a question of whether or not all modal properties are admissible to laws of nature than a question of which, or which kind of, modal properties can be said to be actually possessed by an object. To the latter question, I do not even know where to begin to address it, while to the former question, I suspect that it is the same kind of question as which (categorical) properties are admissible to laws. (Ibid.)

If the question of Bill's possible lawyerliness is a question of which modal property is admissible to laws of nature, and if that question is in essence not any different from the question which *categorical* properties are admissible to laws of nature, then—given that no one is suggesting that it is philosophy's business to tell science which occurrent, i.e. presumably unproblematic, properties it ought to work with—the question of Bill's lawyerliness ought to be decided by science. Basically, the idea is the same as Carnap's for the match: in order to find out about the potentialities of Bill, we would have to find out what he is "made of." If we find that other people who are made of the same stuff as Bill have shown sentimentality and lawyerliness under different circumstances, then we are entitled to say of Bill that he is sentimental and lawyerly as well—his disposition is just being permanently inhibited by contingent factors. In order to find out about Bill's lawyerliness, then, we would have to attempt to (1) formulate the relevant laws linking lawyerliness to other features of Bill that are observable even when lawyerliness does not display, and (2) observe Bill to see whether satisfies the antecedent conditions of those laws. Given that Liu's answer is the same as Carnap's, we can expect him to have problems with the same case as Carnap: what if we want to ascribe a disposition that has never displayed, by no particular of any substance whatsoever? Carnap said that the ascription was meaningless in that case, and was criticized for it by realists. Yet, do they have anything better to offer?

We observe some degree of consistency between fellow realists in the way they handle dispositions: exactly like Mellor, Liu argues that whether or not it makes sense to say of someone that he has disposition X, because there might be favourable conditions under which he would (could) display X, is a question of whether X can figure in scientific laws. This is a move made by all (scientific) realists within the scope of this survey:[150] Fodor, for instance, stipulated that the relevant *ceteris paribus* law containing mental dispositional state X is true (and hence that X was to be allowed into our ontology), if and only if there is a preexisting set of established scientific laws in the antecedents of which X already figures. So did, in their own way, Millikan and Mumford. Unfortunately, it seems quintessential to this sort of approach

[150] Nothing about this sensible move is, in itself, particularly realist. As we have seen, to defer the question of unmanifesting dispositions to science and scientific laws is also something that an empiricist like Carnap or Goodman would recommend, as we have seen in Sec. 2.1. What makes this move realist is a certain interpretation of law-statements: to someone like Goodman, laws just correlate "manifest" predicates that apply only to 'actual things' (which, to his mind, excludes dispositions), whereas a robust realist thinks natural laws contain terms referring to *actually existing* dispositional states.

to give short shrift to two facts: (a) there is a *severe shortage* of laws that we could appeal to in order to decide those dubious cases that move us, and (b), science itself is frankly unconcerned with giving philosophers a helping hand: there is no scientific taxonomy on the sort of ideal conditions it is permissible or plausible to invoke when evaluating disposition-ascriptions or, for that matter, CP-laws (many scientists would argue that there shouldn't/couldn't be any taxonomy, either). Yet, as we shall see presently, this taxonomy is indispensible.

Concerning (a), consider that the question, in Bill's case, does not turn on whether and when we will have accepted laws about sentimentality and lawyerliness, in the sense of confirmed generalizations about the conditions under which these dispositions manifest. What's interesting about Bill is precisely that, by hypothesis, he never manifests sentimentality nor lawyerliness, and he never will, yet he *could* in just the right conditions. So, a simple law defining lawyerliness for us by describing stimulus-manifestation pairs—á la Carnap's reduction sentences—would not do the trick. What we would need is much deeper knowledge about lawyerliness comparable to our knowledge about what sort of chemical structure makes different substances water-soluble. What's the *stuff* of lawyerliness? As Cartwright would put it, does Bill have a lawyerly *nature*? Only this sort of knowledge would allow us to know about what he would do across different conditions (We could have some successful "phenomenological" laws predicting the onset of sentimentality and lawyerliness in people, but without this deeper kind of knowledge of their deeper dispositions, capacities, or natures, we would not have a clue about their *possible* sentimentality and *possible* lawyerliness).

Generally, it seems that inordinately more knowledge is necessary to know whether and when a modal property obtains than when a categorical one obtains. I know how to ascertain whether someone is a good football player if he currently manifests this capacity—by watching him play, for example. But it takes inordinately more experience to ascertain that someone, who is currently awful, could be a good football player. What the experience is needed for is an estimate of the sort of conditions it would require to make his disposition to be a good football player come out into the open. How many years of training, how much psychological support? If these conditions involve, say, transplanting a stronger heart, most of us, and certainly an ordinary coach, would refrain from ascribing the disposition to him. Mumford 1998 concludes, as we have seen, that disposition-ascriptions are context-relative, and would consider the case of the coach to be a case in point. But how convincing is it to suggest that the good-footballer-ascription would be ruled out in the context of a football-pitch, but might possibly stand up in the context of two (granted, very advanced) surgeons discussing whether they can make this patient into a football star? Surely, disposition-ascriptions cannot be truly indexical, varying their truth-value from case to case?

The difficulty increases *dramatically*, as it did for Carnap, if we wish to ascribe properties and dispositions which are even “more” modal than Bill’s lawyerliness. After all, there are lawyers in this world who behave lawyerly, and the problem is merely a fascinating psychological puzzle, namely of establishing laws that would identify lawyerliness on the basis of evidence that does not include the recognized manifestations of lawyerliness. This project, whether realizable or not, seems perfectly sensible and empirical, and so does the question asking whether Bill could have been a lawyer. Even the question whether it is possible to turn this person into a football player (opera singer, scuba diver, what-have-you) still rings somewhat empirical, given that there are football players, opera singers, and scuba divers. All hell brakes loose, however, as we have seen in previous Sections, if we want to ascribe things like omniscience, and fabulous running or flying capacities. Carnap had to capitulate here, saying that given that there is no physically possible test condition for these “dispositions”, it would be nonsensical to ascribe them to anyone (these dispositions are defined as manifesting under optimized conditions—the test would consist in bringing about the optimized conditions, which, by hypothesis, is physically impossible). Was Carnap wrong, and science (and hence philosophy) does have something more to say about cases such as these, beyond throwing its hands up in Wittgensteinian fashion and saying that the bounds of sense have been transgressed?

It seems that in order to be serious, i.e. scientific, about modal properties, we need a *scientific* way of making the sort of estimate the coach, or the surgeon, or even Achilles as he gauges his chances against the turtle, is making. If we refrain from a serious attempt at defining the conditions under which we are willing to ascribe a disposition, then everybody will wind up having every disposition conceivable—or, rather, whether they do will “depend” on the conversational context. That would be contrary to the very *rationale* of ascribing deep dispositions, capacities, or natures: which capacities a thing has is supposed to be a determinant of *what it is*, in an ontological sense, and this is incompatible with it having infinitely many dispositions, or different dispositions according to context. Bill, so the idea, has the stuff of a lawyer, although this is currently suppressed by his banker’s lifestyle. The mode of thinking here is eminently ontological. We don’t want a theory on which he is also a great military commander, actor, playwright, criminal, athlete, philanthropist, etc., *ad infi-*

nitum, although it cannot be denied that he could also have been all these things if we manipulated conditions enough, and chose to make the relevant ascriptions in the appropriate context.

Realists about dispositions have reacted against what they perceived as the exaggerated aversion to modal properties of empiricism by doing a good and healthy thing for the philosophy of science: recognize the undeniable and admit *some* modal concepts and properties at least. To continue to claim, as science progresses, that fundamental concepts, such as force, mass, and charge, cannot denote modal properties, and must turn out to be reducible to categorical properties, seems progressively unreasonable. Yet, this rehabilitation of the ontological status of modal properties threatens to throw open the flood gates for all sorts dubious entities, like my omniscience, or Bill's lawyerliness, which seem to threaten all disposition ascriptions with vacuity. Realists need to partially close the gates, so as to let through only what's desired.

But that's exactly what we are doing!, the realists say. We make sure disposition ascriptions are not vacuous, by admitting as real only those dispositional properties science decides it needs. This leads me to my worry (b). Science—or perhaps I should say scientists with an interest in making their activity epistemologically plausible—is not at all concerned with actually doing the job it has been assigned by philosophers such as Mellor, Fodor, Liu (and Cartwright, as we shall see). Rightly so, as I will argue, because science, quite obviously, cannot do the job realists apparently want it to do. Before I state the argument, a few preliminary remarks on modal properties. Science (or should I say philosophers of science?) has undeniable difficulties with certain modal properties—this is why the empiricists wanted to ban them in the first place. For instance, there are no statistical correlations between people who could be smokers and cancer, only between people who are smokers and cancer. This is because there *cannot* be any statistical connection between people who could be smokers and incidences of cancer, for *anyone* could be a smoker and the property therefore does not discriminate between the target group for cancer. Similarly, the law of Free Fall does not govern the behaviour of *this* man, who could jump from a bridge one day, it only governs the behaviour of those who do jump from a bridge, etc. We could all jump from a bridge if the conditions were right, therefore the extension of '◊ jumps from a bridge' is (at least) all of humanity. Just like in the case of '◊ smokes', if some law does happen to apply to someone who satisfies the predicate '◊ jumps from a bridge', chances are that the predicate does not actually figure in that law, but that the law applies in virtue of some other, perhaps non-modal property. It is undeniable that if a person satisfies the predicate '◊ jumps from a bridge', then various psychological laws would be true of that person that couldn't be true of other things—rocks or washing mashines, for instance, can only *fall* from bridges, not jump from them. But these psychological laws would be true of that person not in virtue of the particular modal property ◊ jumps from a bridge, but rather in virtue of his or her *actual* property of being a human, or a sentient being, or whatever. The same, it seems, goes for all other modal properties.

Or does it? The case is of course not as clear-cut as I have made it out here, otherwise we could all simply return to a strictly empiricist philosophy. True, the scientific value of a modal property like ◊ jumps from a bridge might turn out be rather limited due to the fact that although the property does allow us to discriminate between inanimate objects incapable of intentional action and objects capable thereof, other properties that are easier to observe do that as well. Our interest is likely to be focused not on the modal property itself, but certain other features of the objects that instantiate it. However, these other features of the object could be modal themselves. All that follows from what we have said about ◊ jumps from a bridge is that it is perhaps not a scientifically very *interesting* modal property. The case of 'suicidal' would already be very different, this is, arguably, a modal property very important to psychiatry and psychology. The question which properties are/are not and ought/ought not be interesting—and which properties ought to be studied by science—cannot, it seems, be legislated upon independently of each case. Or so a scientist would say.

3.3.2 Natural Laws and Modal Properties

Remember, however, that we need to partially close the floodgates. Does scientific practice do that for us? I don't see that it does. Can we extract any principles from carefully scrutinizing scientific practice that would show us how to close them? I don't see that we can. Let me give substance to these claims. We have shown above that empiricists and realists about dispositions agree about what to do with unmanifesting dispositions of a certain kind: if this match has spent its life without ever coming into contact with water, infer that it was water-soluble by establishing a law that connects water-solubility to being made of a certain substance; if Bill has spent the life of a banker, infer that he was lawyerly (could have been a lawyer), by establishing a law that connects lawyerliness to other properties we could check Bill for. Here's another example of how realists about dispositions, or capacities, suggest they are ascribed. Nancy Cartwright gives an example of an everyday ascription of what she calls 'capacities', and describes what distinguishes it from a scientific one:

> Contrast two more everyday ascriptions. I am irritable and my husband is inaccurate. These are undoubtedly capacities we have. Ask the children or anyone we work with. Each has been established on a hundred different occasions in a hundred different ways. Like the Coulomb capacity, these too are highly generic. They give rise to a great variety of different kinds of behaviour; the best description of what they share in common is that they are displays of my irritability or Stuart's inaccuracy. These everyday cases contrast with the scientific examples that I am concerned with in the ways we have available to judge when the capacity obtains and when it does not. No one claims in cases like irritability to point to features which you could identify in some other way, independent of my displays of irritability, that would allow you to determine that I am indeed irritable. Philosophers debate about whether there must be any such features: first, whether there need be any at all in the individual who has the capacity, and second how systematic must be the association between the features and the capacity across individuals. Whatever the answer to these questions about everyday capacities, part of the job of science is to find what systematic connections there are to devise a teachable method for representing them. (Cartwright 1999, p. 54)[151]

[151] The "best description" of what all the relevant instances of behaviour have in common, of what makes them cases of being irritable or inaccurate, is that they are displays of the disposition (or capacity) of *being irritable* or *inaccurate*, says Cartwright. She would thus, perhaps, not be disinclined to characterize this situation as one where we fit a certain description onto a series of more or less disconnected events (our evidence): ascribing the disposition as a case of considering the observed displays, seeing the relevant similarity in them, and unifying them by applying the same header. I have argued in Sec. 3.1.2 that this is curve-fitting (and hence of idealization), in so far as the decision whether to apply the description 'irritable' or 'inaccurate' to someone involves (a) considering one's

The key concept is 'the ways we have available to judge when the capacity obtains': what makes the scientific case different from the everyday one is, again, that science is equipped to deal with cases in which the capacity is not actually displaying. We can try to discover 'features which you could identify in some other way, independent of my displays of irritability, that would allow you to determine that I am indeed irritable.' This is what makes science *superior* in its capacity-ascriptions, because thanks to its background knowledge of properties and features usually concomitant with possession of the relevant capacity, science can make ascriptions of capacities that are not actually manifesting. We can tell, by looking at a Tiger that has spent all its life in a Circus cage, certain things it would have the capacity to do if it had the opportunity. This is, we may presume, what uniquely qualifies science to be the ultimate arbiter in matters of modal properties.

Like the other realists, Cartwright actually shares her view of the process by which we ascribe non-manifesting capacities with the empiricists.[152] Does this close the floodgates against my omniscience, or the turtle's running capacity? Cartwright says, in the passage quoted, that the question whether there must be any features in the individual concomitant with possession of the relevant capacity in order for us to be justified to ascribe the features to the individual, and how systematic the connection between features and capacity must be are matters *philosophical*—and these were indeed exactly the sort of issues Fodor 1991 grappled with—, and that 'whatever the answer to these questions about everyday capacities, part of the job of science is to find what systematic connections there are'. She sticks to this line when

---

total evidence about the behaviour of that person, and (b) *reasoning* about this evidence in a way which is isomorphic with the reasoning we deploy in curve-fitting. What is the simplest hypothesis, given our evidence of "hundreds of different occasions"? E.g. is he inaccurate, or is he inaccurate on mornings and evenings; what about apparent counterexamples? Are they anomalies, or do they indicate additional factors, for example the possibility that he is only inaccurate if in a certain mood? Or, alternatively, the possibility that the capacity 'inaccuracy' can, on some occasions, be overridden in its effects by other capacities. Finally, does the description fit with my other background knowledge of him?, etc.

An advantage of construing disposition-ascriptions as curve-fitting in this sense is that this minimizes the difference between the everyday and the scientific case. (Granted, only for cases where we have direct evidence, i.e. a set of manifestations, over which we can fit a curve).

[152] Compare, this time, with Goodman: Goodman, recall, noted that if we wanted to apply a dispositional predicate to an object to which we could not apply any of the manifest predicates associated with the dispositional predicate, then we had the option of looking for law-like connections between manifested predicates that would nevertheless allow us to ascribe the disposition; moreover, he acknowledged that we may not always be able to find many such laws and retracted by saying that "abundant information" may sometimes be sufficient (see Sec. 2.1.2). Apart from terminology, this seems identical to Cartwright. As we shall see, Goodman's shortcomings seem to be also Cartwright's.

pressed, specifically, about the spurious claim that I can fly (see Sec. 2.5.2, *supra*): this is an *empirical* question, she insists, to be decided by investigating whether I display any other features usually associated with the ability to fly. Yet the ability to fly is, of course, a diposition actually displayed by countless individuals, and the question is to be handled according to the known recipe.

Omniscience, on the other hand, or my ability to jump from here to the moon, or to add indefinitely large numbers, are necessarily non-manifesting "capacities", and we can tell, without moving from our philosophical armchair and with complete confidence, that any search for systematic connections between possession of the modal property 'omniscience' and actual properties—even other modal properties 'antecedently declared OK'—is bound to be rather fruitless. Come to think of it, if it were not, we would have an empirical way of searching for God. Now, is the right way to close the floodgates to be "Russellian", and to say that given that there are no features in the actual world usually associated with possession of the relevant capacity that science could investigate, then the ascription is not just meaningless, but *false*? That would seem to be the equivalent of the realist shooting himself or herself in the foot: for the mission statement, ever since our rejection of Carnap's reduction sentences, was to devise a theory that would allow us to be non-verificationists, even realists, about OK modal properties such as my disposition to add.

Here we come to the crux of the matter: the only proper solution of this quintessentially *philosophical* problem, in the present author's view, must be a theory of idealization or idealized laws that does not, á la Fodor and Liu, place the most difficult part into the scientist's lap. It is not science's job to rule out my "omniscience", and the millions of other spurious dispositions that I or a Sceptic could ascribe to me, for doing this would throw out the baby with the bath, and rule out some properties we might wish to keep. In fact, even if there are *no* systematic, i.e. law-like, connections between possession of a modal property and other properties associated with it that we can observe and measure, this does not mean that the capacity does not exist. Let me substantiate this claim. My argument begins like this

> (1) there are infinitely many modal truths about me (in the sense of propositions that are not true of me, but could have been).
> (2) If proposition *p* is a modal truth about A, then A instantiates a corresponding modal property P.
>
> ———————
>
> ∴ I instantiate infinitely many modal properties

I take this much to be fairly uncontroversial, for it depends only on the basic acknowledgements that (1) "things could have been otherwise", and that (2) properties are plentiful. Concerning the first, it seems clear to me that if I was born on January 30th, 1970, at 3:00am sharp, then there are infinitely many propositions about me of the form 'I was born on January 30th, 1970, at 3:00am + X' that could have been true of me (just make X infinitesimally small). Concerning the second, it seems clear that if it is true of me that I was born on January 30th, 1970, at 3:00am, then I instantiate the property (whether "complex" or not) 'was born on January 30th, 1970, at 3:00 am'. There are, of course, other construals of properties where properties are not infinitely many—Armstrong's "sparse" properties are one example. However, there is a clear and established notion of 'property' in philosophy with which I am working here, according to which (2) is true. Whether all these properties can be "natural" or "scientific" is exactly what we are investigating here. Now, the argument continues:

(4) I instantiate infinitely many modal properties

(5) Scientific Realism (Mellor, Liu) about modal properties: Whether a modal property is ontologically real (actually instantiated by me) depends on whether it can figure in scientific laws, and whether I satisfy these laws.[153]

(6) There cannot be infinitely many scientific laws

∴ (7) I do not instantiate infinitely many modal properties.

Contradiction

(6) I assume, is likely to arouse suspicion. The reasoning behind it is that laws, as unanimously acknowledged, are general statements, generalizations of some such form as 'All ... are ...'. If we allowed that a complete science, namely one that accounts for all modal properties, needs to contain infinitely many laws, then (apart from obvious pragmatic things to say about the reduced usefulness of a science with infinitely many laws for finite cognitive agents: no predictions could be made, given that one could never be sure whether one has taken into account all relevant laws. The

[153] Cartwright, who does not have a concept of laws of nature according to which laws of nature (i.e. regularities) refer to modal properties or dispositions, would have something weaker here, involving 'systematic connections'. Whether a capacity is actually possessed by its bearer depends on whether science can observe and measure, and infer the capacity's presence, through some indirect way, independently of manifestations.

situation here is similar to a logical system with infinitely many axioms) it is difficult to see what purpose and advantage such a science would have even for an infinite mind, over simply a list of all true propositions. In order to predict, for example, whether particular *a* is going to be F at time *t* in the future, the infinite mind would simply need to go through the list and attempt to find the true propositions about *a* at *t* and see whether '*a* is F at *t*' is among them. An arduous task, certainly, but no less arduous than the task of going through all laws of nature, and see whether it can derive from them the prediction that *a* is going to be F at *t.* A parallel argument works for explanation. So, in conclusion, it may yet turn out that there are infinitely many true law-like generalizations about nature, and that (6) is false. But if this is the case, then there could not be any human science that pretends to discover these generalizations and deploy them to explanatory and predictive ends. Moreover, even God would have no use for such a science, as she could simply go through her list of true propositions. Importantly for us, if (6) is false, but (5) is true, then no human knowledge of modal properties is possible.

But of course, (5) is false. There is no law for modal properties like '◊jumps from a bridge' and their infinitely many cousins, and—although it is not excluded that for some of them, we will find a true law—for most there never will be. Are we to conclude that, consequently, we instantiate only those very few modal properties science has a law for? Another way of arguing for this conclusion is to reject (1): although it may be correct that there are infinitely many propositions that could be true of me, we ought to reject the notion that I have a corresponding property, disposition, or capacity. Even though there is an extremely remote world at which it is true I could jump from here to the moon, this does not mean that I have the capacity for it. We have argued throughout Chapter 3 that this sort of move would be crippling, ruling out too many properties, dispositions, and capacities that I might reasonably be said to have. Come to think of it, it would too severely hamper our capacity for *thought*, namely for thought about potentiality. Especially if we allow ourselves to think of only those possible worlds that scientific laws provide a model for us.

The upshot of our argument, I take it, is that our reliance on laws must be ended. In a way, this is a conclusion in Cartwright's sense. After all, she wants to point science in a direction where the importance of knowledge of laws is diminished, and the knowledge of capacities or natures, and their modes of combination, is paramount. Yet, Cartwright is also a self-avowed empiricist/"local realist", which means in our context that she will only countenance capacities and natures that science can investigate through its usual methods—Cartwright likes testing. Whether there are any capacities in cases where testing is in principle impossible, is for philosophers to debate about, and hence seems to be of limited interest to her. But how can the question what kind of life Bill could have lead, or whether I am actually adding rather than doing something pathological, be of limited interest? Everyday ascriptions of dispositions, such as “She would have been a great actress”, or even “If he had had a stronger heart, he would have made a formidable rugby player”, are perfectly intelligible, and as such, seem capable of truth value. We can make perfect sense of the question what humans could do in non-nomic situations for which there is little testing, only much theorizing. Philosophers may sometimes be derided for wondering what to do with claims such as ‘If Cesar was in command, he would have used the Atom bomb’. Yet, there are variants of this sort of question that are of the utmost importance and meaning for everyday life: ‘If he wasn’t so melancholic, his marriage would not have broken down’. If his other capacities and natures are such that, were they not accompanied by melancholy, the marriage would have held. All of this, I suppose, scientific laws that rely on establishing connections to occurrent properties will not be able to account for (consider what it would take to establish relevant laws for the last case. For every individual, we would need a tailor-made “law”—but then the knowledge encoded in that “law” would not qualify as lawlike anymore).

The situation is different, of course, with CP-laws, or laws over idealized conditions. This is precisely what they are there for.[154] Yet, we have not found a working methodology for ascribing dispositions of the necessarily unmanifesting kind that would not let in unmanifesting dispositions of the necessarily silly kind—or a methodology for CP-laws in the *absentibus* reading where what’s absent is the world as we know it. One way forward, I have suggested here, is to view disposition-ascriptions, CP-claims, and idealized laws, as being of the same conditional structure, $C \rightarrow (A \rightarrow B)$, and to view the process by which we establish this any claim of this structure as subtended by the same inferential process, manifest in curve-fitting. The dictum ‘curve-fitting is as much art, as it is science’ indicates the sort of problem one encounters when attempting to provide a methodology for curve-fitting, insofar as excluding the relevant grue-like curves seems to involve appeal to a concept of simplicity that escapes formal statistical capture. The only solution is to recognize that

[154] As Lange, M. (2002). “Who's Afraid of Ceteris-Paribus Laws? or: How I Learned to Stop Worrying and Love Them” *Erkenntnis* **57**(407-423) points out, many CP-laws have the advantage that their truth can be preserved under a wide range of counterfactual suppositions, even those that are physically impossible (Lange’s example is that of a CP-law in island biogeography, which would hold even if (some) laws of physics were different; Ibid., p. 418). The same is true for idealized laws.

the world is dappled in a deep way, and that it may contain modal properties, dispositions, and conditional facts not susceptible to yield to standard scientific scrutiny.

This does not mean that they are not susceptible to scrutiny tout court, nor that we cannot attempt to exclude the Sceptic hypotheses. It only means that we may have to accept that additional factors need to be taken into account. These factors are likely to be *pragmatic*.

3.4 Inference to the Best Idealization? Or: Conclusion

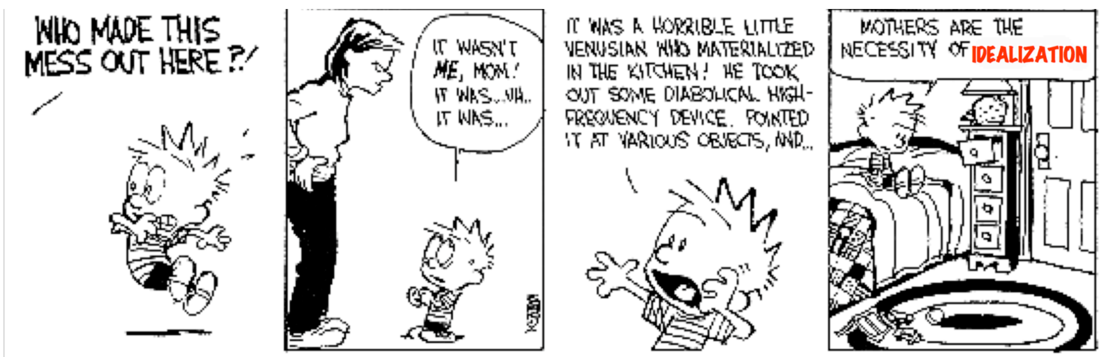

Fig. 17[155]

There is, as our little variation on Calvin and Hobbes suggests, a pragmatic necessity for idealization, and by the same token, for disposition-ascriptions of a certain kind, or CP-claims of a certain form. As Malcolm Forster puts it:

> In every walk of life, and especially in science, people simplify and idealize. For my purposes, an idealization is a model or hypothesis that we *know* to be false. An idealization is *innocuous* if its replacement by a more realistic model does not bring us very much closer to the truth to make the change worthwhile. If it is not innocuous then I will refer to it as ‘harmful’. I claim that the working methodology in almost all intellectual endeavors is to maintain the simplest idealization until it is proven to be harmful. (Forster 2002

[155] Watterson, B. (1992). *The Indispensable Calvin and Hobbes*, Kansas City, Andrews and McMeel, p. 10; used as an illustration by Katarzyna Paprzycka in the “Philosophy of Science”-section of her personal website at http://www.cs.okstate.edu/~marcin/kp/meth.html (accessed 01/2003) In the original, the fourth frame reads of course: “Mothers are the necessity of *invention*”.

As we have seen above, the notion that approximation plays an important role in idealization, or that the concept of 'simplest idealization' can be defined, is dubious. However, Forster is of course right insofar as ubiquity of idealization is concerned, and the fact that the (implicit) working methodology in almost all our cognitive endeavours is to manipulate the sort of idealizations that we, according to some criterion, consider the simplest, or better: *best*. Best for what? Well, for the purpose at hand, seems to be a plausible answer. Explanation, for example, seemed to be Calvin's primary purpose in Fig. 17, and invoking the (possible) actions of a Venusian seemed to fit that purpose best... I have, in places, referred to the sort of inference at work in positing dispositions, fitting curves, or postulating generalizations defined over ideal conditions, as ampliative inference. There is a good philosophical case for arguing that any ampliative inference is a case of inference to the best explanation. This conclusion would certainly seem to be warranted by much of what we have said here. To make it more specific to our problem, we could modify the suggestion and say that the sort of inferences we were concerned with were all inferences to the best idealization.

Glymour 2002 suggests that the best methodology for ceteris paribus claims is a pragmatic one. The same, so it seems to us, is likely to be true for a methodology for the best practice in inference to the best idealization... For the Sceptical problem is a problem that will loom on the horizon of any such methodology, and experience seems to show that pragmatic arguments have traditionally always been the strongest against epistemological scepticism. But this is not the place to further develop this. We have not, I fear, shown the fly the way out of the bottle—perhaps we have been able map out some of the terrain, which will always be the first condition for finding the way out.

In the first Chapter, I had argued that both Kripke's and Goodman's problem could be understood as problems of multiple redescription, and that the relevant sceptical challenge was a challenge to provide factual grounds for the particular choice of description we usually favour. The choice of description or predicate, I noted, is tantamount to a choice of a curve over a set of data likely to be determined by our own internal constraints on the space of possibilites. In Chapter 2, I went on to examine 3 explicit solutions to Kripke's paradox in terms of dispositions, taking issue with the fact that all authors concerned rather neglected epistemological issues: the entities appealed to in their respective solutions are, by no means, observational, and their existence must be inferred on the basis of observed entities or events. The relevant sceptical challenge concerns therefore, as in Chapter 1, the factual basis on which these inferences are made, and the constraints, if any, operating on them. Chapter 3, finally, argued that disposition ascriptions contain substantial elements of idealization. To ward off the danger of vacuity, dispositional theories need to specify limits on legitimate forms of idealization. Disposition ascriptions are, in fact, forms of implicit curve-fitting, I argue, curve-fitting in which our "data" is not necessarily numeric, and the "curve" fitted not necessarily graphic. Nevertheless, the same process is at work. This closes the circle: Goodman's and Kripke's problems are problems concerning curve-fitting, and the solutions for it appeal to entities the postulation of which is the result of curve-fitting. But this does not yet show us the solution, which must come from a methodology governing the sort of idealizations, or inferences to the best idealization, that are a part of curve-fitting. I offered an argument for why natural science could not be expected to be of much help in this domain, given the ubiquity of idealization. If anything, I hope to have pointed towards a different sort of reason for thinking that Wittgenstein was right when he wrote, in the *Tractatus*, that when all scientific questions are answered, our most important and pressing problems will yet have to be addressed.

Index

Bibliography

Allen, B. (1989). “Gruesome Arithmetic: Kripke's Sceptic Replies” *Dialogue* **28**(2): 257-264

Amores, J. G. (2002). “*Morfosintaxis Inglesa.* Unit 2: From Taxonomic to Generative Grammar”, http://fing.cica.es/~gaby/Docencia/Morfo301/Morfo301.htm. (accessed 25/04/2003)

Anscombe, E. (1985). “Wittgenstein on Rules and Private Language” *Ethics* **95**: 342-52

Aristotle (1974). *Categories; and, De interpretatione. Translated [from the Greek] with notes by J.L. Ackrill*, Oxford, Clarendon Press

Aristotle (1997). *Topics: Books I and VII with excerpts from related texts. Translated with a commentary by Robin Smith*, Oxford, Clarendon Press

Armstrong, D. M., C. B. Martin, et al. (1996). *Dispositions: A Debate*, New York, Routledge

Baker, G. P. and P. M. S. Hacker (1984a). “On Misunderstanding Wittgenstein: Kripke's Private Language Argument” *Synthese* **58**: 407-50

Baker, G. P. and P. M. S. Hacker (1984b). *Scepticism, Rules and Language*, Oxford, Blackwell

Barker, S. F. and P. Achinstein (1960). “On the New Riddle of Induction” *Philosophical Review* **69**: 511-522

Bird, A. (1998). “Dispositions and Antidotes” *Philosophical Quarterly* **48**(191): 227-234

Bird, A. (2001). “Necessarily, Salt Dissolves in Water” *Analysis* **61**(4): 267-274

Blackburn, S. (1969). “Goodman’s Paradox” *Studies in the Philosophy of Science*. N. Rescher, Oxford, Blackwell, **3:** 128-42

Blackburn, S. (1984a). “The Individual Strikes Back” *Synthese* **58**: 281-302

Blackburn, S. (1984b). *Spreading the Word: Groundings in the Philosophy of Language*, Oxford, Clarendon Press

Bogen, J. and J. Woodward (1988). “Saving the Phenomena” *Philosophical Review* **97**: 303-352

Boghossian, P. A. (1989). “The Rule-Following Considerations” *Mind* **98**: 507-549

Boghossian, P. A. (1994). “Sense, Reference and Rule-Following” *Philosophy and Phenomenological Research* **54**(1): 139-144

Bowie, A. (1999). “The Meaning of the Hermeneutic Tradition in Contemporary Philosophy” *German Philosophy Since Kant*. A. O'Hear, London

Busemeyer, J. R. and Y.-M. Wang (2000). “Model comparisons and model selections based on generalization test methodology” *Journal of Mathematical Psychology* **44**: 171-189

Carnap, R. (1936-37). “Testability and Meaning” *Philosophy of Science* **3-4**

Carnap, R. (1945). “On Inductive Logic” *Philosophy of Science* **12**: 72-97

Carnap, R. (1947). “On the Application of Inductive Logic” *Philosophy and Phenomenological Research*: 8 133-148

Carnap, R. (1953). “Testability and Meaning” *Readings in the Philosophy of Science*. H. Feigl and M. Brodbeck, New York, Appleton-Century-Crofts

Carnap, R. (1956). “The Methodological Character of Theoretical Concepts” *International Encyclopedia of Unified Science*, Chicago, University of Chicago Press, **1**

Cartwright, N. (1983). *How the Laws of Physics Lie*, Oxford, Clarendon Press

Cartwright, N. (1989). *Nature's Capacities and their Measurement*, New York, Clarendon Oxford Press

Cartwright, N. (1997). “Where Do Laws of Nature Come From?” *Dialectica*: 51(1) 65-78

Cartwright, N. (1999). *The Dappled World. A Study of the Boundaries of Science*, Cambridge, Cambridge University Press

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*, Ny, Praeger

Collett and Land (1978). “How Hoverflies Compute Interception Courses” *Journal of Comparative Physiology* **125**: 191-204
Collins, A. W. (1992). “On the Paradox Kripke Finds in Wittgenstein” *Midwest Studies in Philosophy* **17**: 74-88
Cowie, F. (1999). *What's within? Nativism reconsidered*, New York, Oxford University Press
Davidson, D. (1970). “Mental Events” *Experience and Theory*. J. W. Swanson, The University of Massachusetts Press; Duckworth
Dennett, D. C. (1991). “Real Patterns” *Journal of Philosophy* **88**(1): 27-51
DeVito, S. (1997). “A Gruesome Problem for the Curve-Fitting Solution” *British Journal for the Philosophy of Science*: 48(3) 391-396
Earman, J. and J. Roberts (1999). “"Ceteris Paribus", There Is No Problem of Provisos” *Synthese* **118**(3): 439-478
Earman, J., J. Roberts, et al. (2002). “*Ceteris Paribus* Lost” *Erkenntnis* **57**: 281-301
Essler, W. K. (1970). “An Inductive Solution of the Problem of Dispositional Predicates” *Ratio* **12**: 108-115
Fodor, J. A. (1987). *Psychosemantics*, Cambridge, Massachusetts, MIT Press
Fodor, J. A. (1990). “A Theory of Content II” *A Theory of Content and other Essays*, Cambridge, Masachusetts, MIT Press
Fodor, J. A. (1991). “You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanation” *Mind*: 19-34
Fodor, J. A. (1993). *The Elm and the Expert: Mentalese and Its Semantics*, Cambridge, MIT Pr
Fodor, J. A. (2001). “Doing without What's Within: Fiona Cowie's Critique of Nativism” *Mind* **110**(437): 99-148
Forbes, G. (1984). “Scepticism and Semantic Knowledge” *Proceedings of the Aristotelian Society* **84**: 223-240
Forster, M. (2002). “The Meaning of Temperature and Entropy” *University of Wisconsin*, http://philosophy.wisc.edu/forster/Book/Khinchin.htm. (accessed 12/03/2003)
Forster, M. and E. Sober (1994). “How to Tell when Simpler, More Unified, or Less "Ad Hoc" Theories Will Provide More Accurate Predictions” *British Journal for the Philosophy of Science* **45**(1): 1-35
Forster, M. R. (1999). “Model Selection in Science: The Problem of Language Variance” *British Journal for the Philosophy of Science* **50**(1): 83-102
Forster, M. R. (2000). “Key Concepts in Model Selection: Performance and Generalizability” *Journal of Mathematical Psychology* **44**: 205-231
Franks, B. (1995). “On Explanation in the Cognitive Sciences: Competence, Idealization, and the Failure of the Classical Cascade” *British Journal for the Philosophy of Science*: 46(4) 475-502
Franks, B. (1999). “Idealizations, Competence and Explanation: A Response to Patterson” *British Journal for the Philosophy of Science*: 50(4) 735-746
Frege, G. (1892). “Über Sinn und Bedeutung” *Zeitschrift für Philosophie und philosophische Kritik* **100**: 25–50
Frege, G. (1994). “Funktion und Begriff” *Funktion, Begriff, Bedeutung. Fünf logische Studien*. G. Frege, Göttingen, Vandenhoeck & Ruprecht
Glymour, C. (2002). “A Semantics and Methodology for *Ceteris Paribus* Hypotheses” *Erkenntnis* **57**: 395-405
Goodman, N. (1946). “A Query on Confirmation” *Journal of Philosophy* **43**: 383-385
Goodman, N. (1949). “On Likeness of Meaning” *Analysis* **Vol. 10**: 1–7
Goodman, N. (1953). “On Some Differences about Meaning” *Analysis* **Vol. 13**: 90–96
Goodman, N. (1954). *Fact, Fiction, and Forecast*, London, Athlone Press
Goodman, N. (1970). “Seven Strictures on Similarity” *Experience and Theory*. J. W. Swanson, Boston, University of Massachusetts Press
Goodman, N. (1972). *Problems and Projects*, Indianapolis, Bobbs Merrill

Goodman, N. (1976). *Languages of Art. An Approach to a Theory of Symbols*, Indianapolis, Hackett

Goodman, N. (1978). *Ways of Worldmaking*, Hassocks, Harvester Pr

Goodman, N. (1983). *Fact, Fiction, and Forecast*, Cambridge, Harvard University Press

Hacking, I. (1975). *Why Does Language Matter to Philosophy?*, Cambridge University Press, Cambridge

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge, Cambridge University Press

Handfield, T. (2001). "Dispositional Essentialism and the Possibility of a Law-Abiding Miracle" *Philosophical Quarterly* **51**(205): 484-494

Harman, G. (1994). "Simplicity as a Pragmatic Criterion for Deciding what Hypotheses to Take Seriously" *Grue! The New Riddle of Induction*. D. F. Stalker, La Salle, Illinois, Open Court**:** 153–71

Harnad, S. (1987). "Categorical Perception: The Groundwork of Cognition" *Categorical Perception: The Groundwork of Cognition*. S. Harnad, New York, Cambridge University Press

Hempel, C. G. (1945). "Studies in the Logic on Confirmation, Part I" *Mind* **54**: 1-26

Hempel, C. G. (1966). *Philosophy of Natural Science*, Englewood, Cliffs Nj Prentice Hall

Hempel, C. G. (1988). "Provisoes: A Problem Concerning the Inferential Function of Scientific Theories" *Erkenntnis* **28**: 147-164

Hesse, M. (1969). "Ramifications of 'Grue'" *British Journal for the Philosophy of Science* **20**: 13-25

Horwich, P. (1990). "Wittgenstein and Kripke on the Nature of Meaning" *Mind and Language* **5**(2): 105-121

Horwich, P. (1995). "Meaning, Use and Truth" *Mind* **104**(414): 355-368

Horwich, P. (1998). *Meaning*, Oxford, Clarendon

Hume, D. (1777/1975). *Enquiries concerning human understanding and concerning the principles of morals. Reprinted from the posthumous edition of 1777 and edited with introduction, comparative table of contents, and analytical index by L.A. Selby-Bigge*, Oxford, Clarendon

Hüttemann, A. (1996). *Idealisierungen und das Ziel der Physik: Eine Untersuchung zum Realismus, Empirismus und Konstruktivismus in der Wissenschaftstheorie*, Berlin; New York, de Gruyter

Jackman, H. (2000). "Foundationalism, Coherentism and Rule-Following Skepticism"

Jackson, F. (1975). "Grue" *Journal of Philosophy* **72**: 113-131

Jeffrey, R. C. (1966). "Goodman's Query" *Journal of Philosophy* **63**: 281-288

Joseph, G. (1980). "The Many Sciences and the One World" *Journal of Philosophy* **77**: 773-790

Kant, I. (1787/1983). *Kritik der reinen Vernunft (Erster Teil)*, Darmstadt, Wissenschaftliche Buchgesellschaft

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*, Cambridge, Massachusetts, MIT Press

Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language*, Oxford, Blackwell

Kuipers, T. A. F. (1988). "Inductive Analogy by Similarity and Proximity" *Analogical Reasoning*. D. H. Helman, Dordrecht, Kluwer**:** 299-313

Kukla, A. (1996). "The Theory-Observation Distinction" *Philosophical Review*: 105(2) 173-230

Lakatos, I. (1974). "Science and Pseudoscience" *Conceptus* **8**: 5-9

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*, Chicago, Univ of Chicago Pr

Lange, M. (2002). "Who's Afraid of Ceteris-Paribus Laws? or: How I Learned to Stop Worrying and Love Them" *Erkenntnis* **57**(407-423)

Laudan, L. (1981). "A Confutation of Convergent Realism" *Philosophy of Science* **48**: 19-49

Laurence, S. and E. Margolis (2001). "The Poverty of the Stimulus Argument" *British Journal for the Philosophy of Science*: 52(2) 217-276

Laymon, R. (1985). “Idealizations and the Testing of Theories by Experimentation” *Observation, Experiment, and Hypothesis in Modern Physical Science*. P. Achinstein and O. Hannaway, Cambridge, Massachusetts, MIT Press
Laymon, R. (1987). “Using Scott Domains to Replicate the Notions of Approximate and Idealized Data” *Philosophy of Science*: 54 194-221
Laymon, R. (1989). “Applying Idealized Scientific Theories to Engineering” *Synthese* **81**(3): 353-371
Laymon, R. (1998). “Idealizations” *Routledge Encyclopedia of Philosophy Online (http://www.rep.routledge.com/)*. E. Craig, Routledge, Taylor & Francis Group
Lewis, D. (1983). “New Work for a Theory of Universals” *Australasian Journal of Philosophy* **61**: 343-377
Lewis, D. (1997). “Finkish Dispositions” *Philosophical Quarterly* **47**(187): 143-158
Lipton, P. (1999). “All Else Being Equal” *Philosophy* **74**(288): 155-168
Liu, C. (1999). “Approximation, Idealization, and Laws of Nature” *Synthese* **118**(2): 229-256
Liu, C. (2001a). “Approximations, Idealizations, and Models in Statistical Mechanics”, http://philsci-archive.pitt.edu/archive/00000365/. (accessed June 4)
Liu, C. (2001b). “Laws and Models in a Theory of Idealization”, http://philsci-archive.pitt.edu/documents/disk0/00/00/03/63/index.html. (accessed 29/01/2002)
Loar, B. (1985). “Critical Review of Saul Kripke's *Wittgenstein on Rules and Private Language*” *Noûs* **19**
Loewer, B. and G. Rey (1991). *Meaning in Mind: Fodor and his Critics*, Cambridge, Blackwell
Machover, M. (1996). *Set Theory, Logic, and their Limitations*, Cambridge, Cambirdge University Press
Malzkorn, W. (2000). “Realism, Functionalism and the Conditional Analysis of Dispositions” *Philosophical Quarterly* **50**: 452-69
Manktelow, K. I. (1990). *Inference and Understanding: A Philosophical and Psychological Perspective*, New York, Routledge
Margolis, E. and S. Laurence, Eds.) (1999). *Concepts: Core Readings*, Cambridge, Massachusetts, MIT Press
Martin, C. B. (1993). “The Need for Ontology: Some Choices” *Philosophy* **68**(266): 505-522
Martin, C. B. (1994). “Dispositions and Conditionals” *Philosophical Quarterly* **44**(174): 1-8
Martin, C. B. and J. Heil (1997). “Rules and Powers” *Language, Mind, and Ontology*, Cambridge, Blackwell, **12:** 283-311
Martin, R. M. (1990). “It's Not that Easy Being Grue” *Philosophical Quarterly*: 40(160) 299-315
McDowell, J. (1984). “Wittgenstein on Following a Rule” *Synthèse* **Vol. 58**: 326–363
McGinn, C. (1984). *Wittgenstein on Meaning*, Oxford, Blackwell
McMullin, E. (1985). “Galilean Idealization” *Studies in History and Philosophy of Science* **16**: 247-273
Mellor, D. H. (1974). “In Defense of Dispositions” *Philosophical Review* **83**: 157-181
Mellor, D. H. (1995). *The Facts of Causation*, New York, Routledge
Mellor, D. H. (2000). “The Semantics and Ontology of Dispositions” *Mind* **109**(436): 757-780
Mellor, D. H. (2001). “Realistic Metaphysics. An interview with D. H. Mellor by Anna-Sofia Maurin and Johannes Persson” *Theoria* **67**: 4-21
Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Cambridge, Mit Press
Millikan, R. G. (1989). “Biosemantics” *Journal of Philosophy* **86**: 281-297
Millikan, R. G. (1990). “Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox” *Philosophical Review* **99**(3): 323-53
Mott, P. (1992). “Fodor and Ceteris Paribus Laws” *Mind* **101**(402): 335-346
Mulhall, S. (1989). “No Smoke without Fire: The Meaning of Grue” *Philosophical Quarterly* **39**: 166-189
Mumford, S. (1996a). “Conditionals, Functional Essences and Martin on Dispositions” *Philosophical Quarterly* **46**: 86-92

Mumford, S. (1996b). “Virtus Dormitiva, ha, ha, ha” *Philosopher* **84**(2): 12-15
Mumford, S. (1998). *Dispositions*, Oxford, Oxford University Press
Mumford, S. (2001). “Realism and the Conditional Analysis of Dispositions: A Reply to Malzkorn” *Philosophical Quarterly* **51**(204): 375-79
Niiniluoto, I. (1988). “Analogy and Similarity in Scientific Reasoning” *Analogical Reasoning*. D. H. Helman, Dordrecht, Kluwer**:** 271-298
Nordmann, A. (1990). “Persistent Propensities: Portrait of a Familiar Controversy” *Biology and Philosophy*: 379-399
Norton, S. and F. Suppe (2002). “Why Atmospheric Modelling is Good Science” *Columbia Earthscape*, http://www.earthscape.org/p3/mic01/mic01.pdf. (accessed 02/12)
Nowak, L. (1980). *The Structure of Idealization*, Dordrecht, Reidel
Orear, J. (1982). *Physik*, München, Carl Hanser
Pap, A. (1963). “Reduction Sentences and Disposition Concepts” *The Philosophy of Rudolf Carnap*. P. A. Schillp, La Salle, Illinois
Patterson, S. (1998). “Competence and the Classical Cascade: A Reply to Franks” *British Journal for the Philosophy of Science* **49**(4): 625-636
Peacocke, C. (1992). *A Study of Concepts*, Cambridge, MIT Pr
Pears, D. (1988). *The False Prison: A Study of the Development of Wittgenstein's Philosophy (Vol. 2)*, New York, Clarendon
Peirce, C. S. (1931-35). *Collected Papers (Vols. I-VI)*, Cambridge, Massachusetts, Harvard University Press
Pietroski, P. and G. Rey (1995). “When Other Things Aren't Equal: Saving "Ceteris Paribus" Laws from Vacuity” *British Journal for the Philosophy of Science* **46**(1): 81-110
Priest, G. (1976). “Discussion: Gruesome Simplicity” *Philosophy of Science*: 43 432-437
Prinz, J. J. (2001). “Philosophy of Cognitive Science: Chomsky’s Linguistics”, http://www.artsci.wustl.edu/~jprinz/cogsci2001/cog2001-3.htm. (accessed 25/4/03)
Puhl, K., (Ed. (1991). *Meaning Scepticism*, Berlin, de Gruyter
Putnam, H. (1980). “Models and Reality” *Journal of Symbolic Logic* **45**: 464–482
Quine, W. V. O. (1960). *Word and Object*, Cambridge, MIT Press
Quine, W. V. O. (1969). *Ontological Relativity and Other Essays*, New York, Columbia University Press
Ray, G. (1997). “Fodor and the Inscrutability Problem” *Mind and Language* **12**(3-4): 475-489
Rosch, E. (1977). “Linguistic Relativity” *Thinking*. P. C. Wason, Cambridge, Cambridge University Press
Rosch, E. and C. Mervis (1975). “Family Resemblances: Studies in the Internal Structure of Categories” *Cognitive Psychology* **7**
Rubinstein, A. (1998). “Induction, Grue Emeralds and Lady Macbeth's Fallacy” *Philosophical Quarterly* **48**(190): 37-49
Russell, S. (1988). “Analogy by Similarity” *Analogical Reasoning*. D. H. Helman, Dordrecht, Kluwer**:** 251-269
Ryle, G. (1949). *The Concept of Mind*, London, Hutchinson's University Library
Sainsbury, R. M. (1995). *Paradoxes*, New York, Cambridge University Press
Savigny, E. V. (1988). *Wittgensteins "Philosophische Untersuchungen": Ein Kommentar für Leser (Band 1)*, Frankfurt, Klostermann
Schiffer, S. (1991). “*Ceteris Paribus* Laws” *Mind* **100**: 1-17
Schurz, G. (2001). “Pietroski and Rey on Ceteris Paribus Laws” *British Journal for the Philosophy of Science* **52**(2): 359-370
Shoemaker, S. (1975). “On Projecting the Unprojectible” *Philosophical Review* **84**: 178-219
Skyrms, B. (1993). “Analogy by Similarity in Hyper Carnapian Inductive Logic” *Philosophical Problems of the Internal and External Worlds*, Pittsburgh, University of Pittsburgh Press
Sober, E. (1982). “Dispositions and Subjunctive Conditionals, or, Dormative Virtues Are No Laughing Matter” *Philosophical Review* **91**: 591-596

Sober, E. (1994). “No Model, No Inference: A Bayesian Primer on the Grue Problem” *Grue! The New Riddle of Induction*. D. Stalker, Chicago, Open Court
Stegmüller, W. (1989). *Hauptströmungen der Gegenwartsphilosophie (Vol. 4)*, Stuttgart, Alfred Kröner Verlag
Stroud, B. (1990). “Meaning, Understanding and Translation” *Canadian Journal of Philosophy* **16**: 343-361
Stroud, B. (2000). *Meaning, Understanding, and Practice. Philosophical Essays*, New York, Oxford University Press
Thompson, J. J. (1966). “Grue” *Journal of Philosophy* **63**
Ullian, J. S. (1961). “More on "Grue" and Grue” *Philosophical Review* **70**: 386-389
Watterson, B. (1992). *The Indispensable Calvin and Hobbes*, Kansas City, Andrews and McMeel
Whorff, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. Edited by John B. Carroll*, Cambridge, Massachusetts, MIT Press
Wilson, G. M. (1998). “Semantic Realism and Kripke's Wittgenstein” *Philosophy and Phenomenological Research* **58**(1): 99-122
Wittgenstein, L. (1922). *Tractatus logico-philosophicus*, London, Kegan Paul, Trench, Trubner
Wittgenstein, L. (1967). *Philosophical Investigations. Translated [from the German] by G.E.M. Anscombe*, Oxford, Blackwell
Wittgenstein, L. (1978). *Remarks on the Foundations of Mathematics*, Oxford, Basil Blackwell
Wright, A. (1991). “Dispositions, Anti-Realism and Empiricism” *Proceedings of the Aristotelian Society* **91**: 39-59
Wright, C. (1980). *Wittgenstein on the Foundations of Mathematics*, Cambridge, Harvard University Press
Wright, C. (1997). “The Indeterminacy of Translation” *A Companion to the Philosophy of Language*. C. Wright, Oxford, Blackwell