The Dualist

Stanford's Undergraduate Journal of Philosophy

Winter 2008

For Alexei, who gave us Hilbert Deductions when were overwhelmed

THE DUALIST

Volume XIV ■ Winter 2008

Department of Philosophy
Stanford University

THE DUALIST
Volume XIV
Winter 2008

TABLE OF CONTENTS

DICTIONARIES, DOGMAS, AND ANALYTICITY

PHILIP ATKINS
University of Minnesota Duluth

Abstract

In "Two Dogmas of Empiricism" W. V. Quine attacks the notion of analyticity, which holds that certain expressions are true by virtue of meaning alone. Contra Quine, I argue that semantical rules, understood as conventional stipulations, do provide grounds for analyticity. I show that Quine's argument rests on a weak epistemic consideration, namely, that we cannot know with certainty that such semantical rules obtain. While this may be true, I argue that because we can form strong hypotheses in support of semantical rules, we are well within our epistemic rights in holding that they exist. Dictionary definitions, in my view, are reports of hypothesized semantical rules. In addition, I point out that Quine is inconsistent in arguing that necessity statements (' "Necessarily, all bachelors are unmarried" is true') are equivalent to analytic statements (' "All bachelors are unmarried" is analytic'), since such an argument would require an account of synonymy, which Quine rejects early on.

Introduction

In his 1953 essay "Two Dogmas of Empiricism" W. V. Quine attacks the foundations of both the analytic-synthetic distinction and reductionism.[1] Our concern is only with the former dogma, which maintains that statements are primarily of two separate kinds: *analytic* statements are those that are true by virtue of meaning alone, while *synthetic* statements are those whose truth is dependent upon extra-linguistic fact. Quine believes this to be a central distinction made in philosophy, but ultimately a mistaken one. He rejects the theory of analyticity altogether and since the publication of his paper so have many in the philosophical community. The present essay will outline Quine's basic argument and then critique it while providing a rudimentary account of analyticity. Principally, my account will hinge upon a robust theory of meaning, which itself shall rely on semantical rules. We shall confine our discussion to Quine's treatment of analyticity in "Two Dogmas" and consider nothing more.

Prima facie, it is unclear exactly how much of Quine's argument turns on epistemic considerations. We shall return to this issue later on and I will attempt to show that, in the last analysis, Quine's argument rests upon a weak epistemic worry. In any event, Quine's overall strategy in "Two Dogmas" is clear. Quine essentially launches a two-pronged attack on analyticity. "On the first front Quine argues that the analytic-synthetic distinction has not been clearly drawn; on the second front he argues that it is a mistake to think that the distinction needs drawing."[2] The second front regards reductionism. As previously indicated, our

1 W. V. Quine, "Two Dogmas of Empiricism," in *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953). Quine originally published "Two Dogmas of Empiricism" in the January 1951 issue of the *Philosophical Review*. He reprinted the essay with slight alterations in 1953's *From a Logical Point of View* (Harvard University Press). It is naturally the latter version that is most widely known and it is the latter version that I use for the present essay.

2 Roger F. Gibson, *The Philosophy of W. V. Quine: An Expository Essay* (Tampa, Florida: University Presses of Florida,

critique will involve the first front, our argument being that a boundary between analytic and synthetic statements can be drawn.

Before we proceed it would be best to establish some points on which we agree with Quine. Like Quine, we shall adopt a model which holds that there are two classes of putatively analytic statements: (1) Statements which are *logically true*, and (2) Statements which are *logically true* through *synonymy*.

Statements of the former class are easy to identify. Take as an example, 'All unmarried men are unmarried men'. Such statements are true regardless of any interpretation one gives to 'men' or 'married'. They are, as Leibniz observed, true in all possible worlds, so long as the logical particles ('all', 'un-', 'if', 'and'…) are taken for granted. These statements do not appear to bother Quine,[3] who writes that "the major difficulty lies not in the first class of analytic statements, the logical truths, but rather in the second class, which depends on the notion of synonymy."[4]

Consider a statement from the second class of analytic statements: 'All bachelors are unmarried men'. This is analytic in that it may become a member of the first class (that is, it may become a logical truth) by replacing 'bachelors' with 'unmarried men'. These terms are held to be synonymous and it is through synonymy that the analyticity of the second class consists. It is this notion of synonymy that Quine argues is unclear. We must, therefore, determine adequate grounds for the existence of synonymous terms. We must explain how it is that (2)-class statements reduce to (1)-class logical truths.

Definition

Quine first considers definition as a possible explanation of synonymy and writes, "Certainly the definition which is the lexicographer's report of an observed synonymy cannot be taken as the ground of the synonymy."[5]

For Quine, observed synonymies are supposed synonymies, occasions when two or more terms look like they're being used or regarded as if they meant the same. Thus, observed synonymy is observed linguistic behavior. Here, Quine attempts to expose circularity. His argument is as follows.

i. A dictionary's definition is nothing more than a report of observed synonymy.
ii. The observed synonymy comes from us (i.e., the lexicographer observed it in our use).
iii. Thus, to explain observed synonymy in terms of definition is to explain observed synonymy in terms of observed synonymy.

Of course, (iii) will not do. The dictionary is simply an empirical report of observed (supposed) synonymy and hence it cannot be taken as the basis of that synonymy.

All of Quine's arguments against definition as a means of explaining synonymy are in the same vein: definition, according to Quine, rests or relies upon prior occurrences of supposed synonymy and so cannot be expected to explain those occurrences. Indeed, this is a recurring tactic throughout "Two Dogmas": "Quine's paper… argues that, contrary to appearances, the notion of analyticity cannot be defined or explained without circularity."[6]

1982), 96.

3 In fact, Quine does take issue with (1)-class logical truths as I have described them, i.e., as statements whose truth is given purely by the meanings of a few logical words. Quine set out arguments against this theory of logical truth in both "Truth by Convention" (1936) and "Carnap and Logical Truth" (1954). In "Two Dogmas of Empiricism", however, Quine offers no considerations against class-(1) analyticity. Thus, we shall not pursue the matter any further.

4 Quine, "Two Dogmas of Empiricism," 24.

5 Ibid., 24.

Interchangeability

Another attempt to explain synonymy appeals to the principle of *interchangeability*. Two terms are interchangeable if either may be used in all contexts without a shift in truth value. But this, too, is doomed to failure. The interchangeability may merely reflect sameness of extension, but two terms could always have the same extension and differ in meaning (consider 'creatures with a kidney' and 'creatures with a heart'). So, extensional agreement may be an "accidental matter of fact."[7]

We wish to say that all bachelors are necessarily unmarried men, not incidentally. Of course, if a language does have such a word as 'necessarily', then Quine's challenge to explain the analyticity of 'All bachelors are unmarried men' could be more easily met—one could simply say that 'Necessarily, all and only bachelors are unmarried men'. But here again Quine perceives a problem: it seems that 'necessarily' is only understood insofar as analyticity is already understood. As Quine sees it, ' "Necessarily, all and only bachelors are unmarried men" is true' is equivalent to ' "All and only bachelors are unmarried men" is analytic'. Hence, the latter cannot be explained in terms of the former. Quine remarks, "Our argument is not flatly circular, but something like it. It has the form, figuratively speaking, of a closed curve in space."[8]

Semantical Rules

Some believe that analyticity's unclearness results from the vagueness implicit in ordinary language. Quine disagrees and from hereon considers a more *artificial language* with explicit *semantical rules* in his case against semantical rules.

This move toward semantical rules marks a "second strategy" in explaining analyticity, one that is employed only after the "first strategy" (involving synonymy, definition, and interchangeability) has been found lacking: "Quine's inability to make satisfactory sense of the notion of synonymy … suggests to him that perhaps, instead of trying to clarify synonymy outright (so as to reduce statements like (2) to statements like (1) and thereby to render the concept of analyticity as perspicuous as is the concept of logical truth), he might better try to clarify analyticity outright, and then define 'synonymy' in terms of analyticity."[9] This means that Quine has surrendered (or at least suspended) our basic premise that analyticity consists in the reduction of (2)-class statements to (1)-class statements. Such a departure is, I think, unwarranted. After all, Quine could consistently treat semantical rules as another possible means of explaining synonymy. Indeed, in the critical portion of this essay we shall maintain a two-class model of analyticity and I shall attempt to demonstrate how semantical rules can provide for the reduction of (2)-class statements to (1)-class logical truths.

Before setting out his criticisms of semantical rules, Quine observes that "the notion of analyticity about which we are worrying" is an alleged relation between a statement *S* and a given language *L*.[10] He feels this to be the case for natural lan-

6 Ilham Dilman, *Quine on Ontology, Necessity, and Experience: A Philosophical Critique* (Albany, NY: State Univeristy of New York Press, 1984), 74.

7 Quine, "Two Dogmas of Empiricism," 31.

8 Ibid., 30.

9 Gibson, *The Philosophy of W. V. Quine*, 99.

10 Quine, "Two Dogmas of Empiricism," 33.

guage as well as for artificial. The difficulty lies in explaining the supposed relation between *S* and *L* – that is, in making sense of '*S* is analytic for *L*'. Quine identifies several kinds of semantical rules and offers several criticisms. Unfortunately I have time here to review only one of Quine's arguments. Generally speaking, on the semantical rules approach an analytic statement would be any statement that is not simply true, but *true according to semantical rules*. The problem here, according to Quine, is that we seem to have explained one unclear term ('analytic') in terms of another unclear term ('semantical rule'). Suppose I were to challenge you to explain '*S* is analytic for *L*'. You could conceivably make an appeal to the rules that govern *L*, responding "*S* is analytic for *L* because *S* is true by virtue of the semantical rules of *L*." But you have failed to engage the problem, for one could argue that 'semantical rules of' is in need of as much clarification as 'analytic for'. Thus, the attempt to explain analyticity through semantical rules is as much a failure as any attempt we have so far considered.

Semantical Rules Revisited

Let us begin our critique, first with Quine's contention over semantical rules and working our way back to his discussion of definition. Bear in mind our purpose: to reasonably demonstrate how a (2)-class statement may become a (1)-class logical truth, thus explaining analyticity. Not only shall I attempt to explain analyticity, I shall do so without appealing to artificial language.

Recall that Quine argues that semantical rules fail to provide for analyticity because 'semantical rules of' is in need of as much clarification as 'analytic for'. For Quine, to clarify analyticity in terms of semantical rules would be to trade one unexplained expression for another, ignoring the philosophical problem at hand. To address Quine's worry we need to flesh out the notion of a semantical rule, but how? Quite simply, I argue that semantical rules amount to *conventional stipulations*, generally of the form "Employ linguistic token x only when y." Call y the "employment condition(s)" of x. An easy example of a semantical rule might be

(C) Employ 'the cat is on the mat' only when the cat is on the mat.

Here, the linguistic token is information bearing and the employment condition is a state in the world. In virtue of the way these kinds of semantical rules constrain our behavior, linguistic tokens are capable of indicating such states in the world.[11] The token 'the cat is on the mat', for instance, indicates that the cat is on the mat. The kind of rule relevant to analyticity is, I think, a *sub-rule* of the above kind. An example of a sub-rule would be

(c) Employ 'cat' only when indicating states involving something with the property of being a cat.

Here, the token is an individual word which, when taken in isolation, does not bear information at all. Note how the employment condition concerns an indicating-function—the rule stipulates that 'cat' be employed only when *indicating* cat-states in the world. Thus, because they are parasitic on the indicating-function provided by rules like (C), rules like (c) are semantical sub-rules.[12]

11 David Cole, personal communication.

12 Since declarative sentences occur as the linguistic tokens in semantical rules and individual words occur as the tokens in semantical sub-rules, it may seem that declarative sentences are somehow *prior* to individual words. This may strike some as counterintuitive. Here I shall say only the following in my defense: rather than words being prior to indicative sentences, I prefer to think of words as being largely *dependent* on declarative sentences for their meaning, i.e., words have meaning only insofar as they are used to indicate states in the world. This is not a far cry from the theory that words have sense only in the context of a sentence, which is endemic to the analytic tradition (see Frege, 1884).

Earlier it was said that analyticity consists in the reduction of (2)-class statements to (1)-class logical truths. Indeed, semantical sub-rules provide us with a means of reduction, for they provide us with an account of synonymy. Two terms are synonymous, I submit, when they share the same employment condition(s). Now consider the following example of a semantical rule:

(b1) Employ 'bachelor' only when indicating states involving that which has the property of being male and the property of being unmarried.

Here, the employment condition is the indication of bachelor-states in the world. Now consider:

(b2) Employ 'unmarried man' only when indicating states involving that which has the property of being male and the property of being unmarried.

Although the linguistic token in (b2) is different than that in (b1), the two tokens share identical employment conditions, hence the synonymy relation between 'bachelor' and 'unmarried man'.

Quine himself considered conventional stipulation as a means of clarifying synonymy in his discussion of definition.

> There does, however, remain still an extreme sort of definition which does not hark back to prior synonymies at all: namely, the explicitly conventional introduction of novel notation for the purposes of sheer abbreviation. Here the definiendum becomes synonymous with the definiens simply because it has been created expressly for the purpose of being synonymous with the definiens. Here we have a really transparent case of synonymy created by definition; would that all species of synonymy were as intelligible.[13]

Quine is referring to the kind of "novel notation" that commonly develops in specialized fields. It is common in philosophy, for instance, to substitute 'if and only if' with 'iff'.

Now, let me iterate some of the salient points that spell trouble for Quine. First, there is the acknowledgement that *synonymy occurs*. What's more, it occurs through the kind of conventional stipulations that make up semantical rules and these stipulations altogether avoid circularity. Quine admits as much and this, I think, is a significant step toward analyticity.

Why will Quine allow for synonymy in some cases but not others? Why is Quine willing to believe that a synonymy relation obtains between 'if and only if' and 'iff' but not between 'unmarried man' and 'bachelor'? The answer seems to be that "for terms naturally occurring in ordinary languages there are no *explicit* acts of definition … The very feature which allows us to say of the results of explicitly conventional introduction of novel notation that they are true by definition are *unavailable* in the standard cases in natural language."[14] Since the synonymy that may occur in natural language is not explicit, it is rendered uncertain and therefore unclear.

This is plainly an epistemic worry and not one that carries a lot of weight. Strict certainty concerning semantical rules is a tall order. It is, of course, appropriate to ask how one justifies one's belief in semantical rules. The answer is that through hypothesis we can make strong inferences about the semantical rules operating in language (and thus about synonymy and analyticity). Many of our beliefs lack strict certainty but are justified nonetheless. Imagine that you are exploring a foreign country to gain insight into the strange ways of the native people. You observe a particular practice that occurs repeatedly under similar circumstances and

13 Quine, "Two Dogmas of Empiricism," 26.

14 Richard Creath, "Every Dogma Has Its Day," *Erkenntnis* 35 (1991): 353-54, my italics.

you hypothesize that the practice results from a rule that governs behavior. Those who fail to act in accord with the hypothesized rule are gently admonished and the natives all concur with your general assessment of their culture. It must be admitted that you lack certainty concerning the rule in question, but it seems that you are well within your epistemic rights in holding that such a rule exists. It is exactly this sort of justification that we have with respect to semantical rules. Imagine that someone is intrepidly exploring our own country and hypothesizes that there is a rule such that, for every employment of 'bachelor', such and such a set of conditions must be satisfied. Surely the explorer's hypothesis is a strong one.

These hypotheses are even testable.[15] Suppose that one believes 'bachelor' to be synonymous with 'unmarried man'. To gain further justification, all one need do is go out into the world and ask those who participate in the language under what conditions they employ the word 'bachelor', or if they think that 'All bachelors are unmarried men' is true by virtue of meaning. And, when investigating rules that presumably spring from convention, what better test to conduct than one that openly appeals to the language-users that ultimately determine convention?[16]

Do we have certain a priori knowledge that 'All bachelors are unmarried men' is an analytic statement? Not at all, but this does not mean that the epistemic status of the analytic-synthetic distinction is so precarious as to warrant its rejection. A belief may be justified without being *a priori*, even those beliefs regarding meaning. How this bears upon the putative *a priority* of analytic statements *themselves* is an interesting question, but one which falls beyond our purview. I shall say only the following in passing. While ' "All bachelors are unmarried men" is an analytic statement' is only knowable *a posteriori*, it does not follow that the truth expressed by 'All bachelors are unmarried men' is only knowable *a posteriori*. On the other hand, it is not obviously false that analytic statements are *a posteriori* in character.[17] After all, one must have contact with one's environment to become acquainted with semantical rules. Certainly there are those who lack the requisite contact with their environment. Thus, one might assert that all bachelors are unmarried men and not know that one's assertion is analytic. By the same token, one is not necessarily a logical idiot for asserting that no bachelors are unmarried. Here it may be instructive to draw an analogy between semantical rules and law. Traffic laws, for instance, obtain regardless of whether or not you are obeying the speed limit and they obtain even if you are unaware of their existence. Conversely, traffic laws obtain even if *everyone* is conscientiously obeying them and *no one* is actually setting the laws into action. Exactly how this happens concerns the ontological status of rules, a subject too peculiar and difficult to be suitably addressed in the present essay. I will only remark that the existence of rules seems to rely solely on their power to constrain our behavior, and that the power of rules to constrain our behavior seems to rely solely on the mere *possibility* of their enforcement.[18]

15 Carnap, I discovered, makes a similar case in "Meaning and Synonymy in Natural Language" (1956). There Carnap is concerned to defend the thesis that an ascription of intension is an empirical hypothesis that can be tested by observation of language behavior. This is explicitly contrary to Quine's stance that the ascription of intension is not a matter of fact at all, but merely a matter of choice.

16 In *Semantic Theory* (1972), Katz draws attention to interesting semantic data, such as subjects' agreement over assignments of synonymy, redundancy, implication, etc.

17 The notion of analytic a posteriori truth is not a new one. David Cole argued for it in "Analytic A Posteriori Truth" (2003). Similarly, Saul Kripke (1980) and Hilary Putnam (1975) have both argued for necessary a posteriori truth.

18 This point deserves more qualification. It seems that several counterintuitive conclusions follow from my remarks. For example, if we drive down a street and there are no policemen on patrol, then it would mean that there is no practical possibility of the speed limit being enforced. Thus, according to my remarks, the rule that stipulates the speed limit does not exist, since the (practical) possibility of enforcement is required for a rule's

We are now justified in the belief that there exist semantical rules and that they account for the analyticity of statements such as 'All bachelors are unmarried men'. What's unjustified is the depiction of the analytic-synthetic distinction as "a metaphysical article of faith."[19] To motivate the abandonment of a deeply enrooted and generally consistent philosophical theory, Quine must do more than show that it falls short of limpid certainty.

In sum, we have a means of properly reducing (2)-class statements to (1)-class statements via semantical rules. On this theory, we may say, with respect to any given language *L*, that *S* is an analytic statement for *L* if *S* is either a logical truth or synonymous with a logical truth according to the semantical rules of *L*. Granted, the semantical rule system is itself arbitrary. But the rules, however arbitrary, remain the objective foundation of linguistic meaning—objective insofar as we can speak objectively of rule-following activities. To be sure, the rules are arbitrary and if the rules change so will the class of analytic statements. But this is hardly an objection. The revision of meaning will naturally bear upon the analytic-synthetic distinction (regardless of one's theory of analyticity), but from this possibility it does not follow that such rules cannot accommodate analyticity here and now. We remain justified in our belief that the truth of *S* is a consequence of the semantical rules of *L* and not of any state in the world apart from linguistic forms.

Necessity, Synonymy, and Quine

Quine argued early on that even if a language does include 'necessarily' or 'necessity' as words, it will not help to explain analyticity. The idea is that 'necessarily' is only understood once 'analytic' is understood: "Does ['necessarily'] really make sense? To suppose that it does is to suppose that we have already made satisfactory sense of 'analytic'."[20] Quine's claim is that to say something like ' "Necessarily, all and only bachelors are unmarried" is true' is to say ' "All and only bachelors are unmarried" is analytic'. We may concede this point, but there is still a glaring problem with this line of reasoning when it is examined in light of the general argument presented in "Two Dogmas". Here, the crux of Quine's argument is that a sameness relation exists between statements such as 'Necessarily *p* is true' and '*p* is analytic', so one cannot be explained in terms of the other. Yet we find that Quine contends that there is, in fact, no adequate basis for synonymy, which is why it could not be utilized to clarify analyticity. Quine holds that the notion of synonymy is just as problematic as that of analyticity. "If Quine is to be consistent … then it appears that he must maintain … that the distinction we suppose ourselves to be marking by the use of the expressions 'means the same as,' [and] 'does not mean the same as' does not exist."[21] Hence, Quine cannot consistently argue that synonymy is too unclear to explain analyticity while simultaneously maintaining that necessity cannot explain analyticity because "necessity statements" mean the same as "analytic statements."

One might try to salvage Quine's case, claiming that "necessity statements" merely *presuppose* "analytic statements." So the trouble is not that the one is syn-

existence. This is an odd conclusion: it seems like the rule continues to obtain, only there's no one to enforce it. But where does the rule's existence lie? Not in the thought, 'There might be cops, I might be pulled over'. Surely the rule obtains regardless of what I am thinking. Not in my behavior, for surely the rule obtains regardless of how I act. Besides, it is strange that a rule's existence should have *nothing* to do with its enforcement. I shall only suggest that an optimality condition be introduced. The rule's existence relies upon the possibility of its enforcement "where conditions are optimal" (e.g., if we could have a policeman patrolling every street).

19 Quine, "Two Dogmas of Empiricism," 37.

20 Ibid., 30.

21 H. P. Grice and P. F. Strawson, "In Defense of a Dogma," *Philosophical Review* 65 (1956): 75.

onymous with the other, it's that the one presupposes the other. To be sure, saying '*p* is synonymous with *q*' is not the same as saying '*p* presupposes *q*'. But even on this account of Quine's argument we run into difficulty. Most prominently, '*p* presupposes *q*' seems to be equivalent to '*p* assumes *q*', or better, '*p* only on the assumption that *q*'. But Quine cannot readily allow for these locutions, for they may provide for analyticity. One might argue, for instance, that the truth of those sentences which contain the word 'bachelor' *assumes* or *presupposes* the truth of those sentences which are formed by substituting for 'bachelor' the expression 'unmarried man', and then again that these latter sentences presuppose the former; thus, we may have a basis for analyticity. (As a first pass, we might conjecture that an analytic statement is one which presupposes, and is presupposed by, a logical truth.) If Quine is not opposed to the notion of presupposition, which he is not if he employs the notion himself, then he will have no obvious reason for denying presupposition as a possible basis for analyticity.

Defining Dictionaries

Finally, we arrive at definition. Let us review Quine's points on the subject. Definition cannot adequately explain synonymy because it seems to presuppose prior relations of synonymy. We may recall that Quine's argument is as follows.

i. A dictionary's definition is nothing more than a report of observed synonymy.
ii. The observed synonymy comes from us (i.e., the lexicographer observed it in our use).
iii. Thus, to explain observed synonymy in terms of definition is to explain observed synonymy in terms of observed synonymy.

Presuming the argument's validity, we shall enquire instead into the truth of its premises. It is my contention that (i) is false. A dictionary is not *merely* a report of what people say. Rather, a dictionary might be thought of as a report of a language's semantical rules. The difference is that, on my view, a dictionary can have normative force. A dictionary, I submit, serves as a measure of *semantic correctness* because it records rules.

Quine seems to support (i) by pointing out that lexicographers create dictionaries by going out into the world and observing behavior. "The lexicographer," writes Quine, "is an empirical scientist, whose business is the recording of antecedent facts."[22] What the lexicographer records are instances of supposed synonymy in our linguistic behavior. These observations lead eventually to the formation of a dictionary. The critical point, however, is that because lexicographers observe our behavior, it must then follow that dictionaries are purely reports of observed synonymy and that they cannot be, in *themselves*, stipulative of meaning. Now I take this argument to be invalid, but I shall not argue that here. My contention is not that dictionaries themselves are stipulative of meaning, but that dictionaries are reports of the *semantical rules* that stipulate meaning. We can think of a lexicographer as being something like our intrepid explorer, forming *hypotheses* as to the underlying rules of language. These hypotheses culminate in a report that is authoritative insofar as it is an accurate portrayal of semantical rules. So one may appropriately turn to dictionaries as a measure of semantic correctness, for they contain strong hypotheses concerning the rules that govern language.

This is not, of course, to say that dictionaries are infallible. We can imagine a

22 Quine, "Two Dogmas of Empiricism," 24.

mistaken dictionary, even a radically mistaken dictionary. Dictionaries can be fallible because, on our analysis, rules exist apart from dictionaries. A dictionary is only a measure of correctness because it records rules, not because it constitutes them. In the same way, law books are only authoritative insofar as they accurately record the law. Law, rather than a simple function of what law books say, is a complex social practice that persists with or without law books. Exactly how this occurs is a topic for another essay (the answer concerns the ontology of rules or laws). But it seems right enough that the United States would remain a lawful country even if we were to burn every book on United States law.

We can press even further the analogy between laws and semantical rules: laws, like semantical rules, are subject to revision. Consequently, many law books have fallen into disuse over the years, as have many dictionaries. This occurs when the relevant dictionaries and law books cease to offer an adequate depiction of rules and law.

Lastly, I should like to draw attention to how much reciprocity there is between dictionaries and semantical rules. This reciprocity results from the relation between dictionaries and linguistic conventions. The picture that I endorse is as follows: conventions produce rules, rules comprise dictionaries, and dictionaries, in turn, influence convention.[23] This sort of reciprocity is not surprising, nor is it unique to the relation between dictionaries and semantical rules. In the fashion industry, for instance, aesthetic preferences are largely determined by convention, but then conventions are routinely influenced by the efforts of the fashion industry. Similarly, Hollywood's film ventures are determined largely by the preferences of moviegoers, but a moviegoer's tastes are inevitably affected by the films that Hollywood churns out. Here again there is reciprocity. However apt (or inapt) these analogies are, it is evident that the existence of such a relation between dictionaries and rules does little harm to my picture. Dictionaries espouse definitions; a definition, rather than a mere report of linguistic behavior, is a well-formulated hypothesis concerning the semantical rules that stipulate synonymy.[24]

Conclusion

In the course of our enquiry, we have briefly highlighted Quine's key arguments against analyticity. Our aim was to illuminate the notion of synonymy, which would bridge the gap between class-(2) and class-(1) analytic statements. Despite the influence that "Two Dogmas" has held, we have found no reason to suppose that any of Quine's arguments defeat semantical rules as a possible explanation of analyticity. These rules, understood through hypothesis, may shed light upon the notion of synonymy and revitalize the passé tradition of an analytic-synthetic distinction. Analyticity, from this point of view, pertains to a certain set of statements whose truth is a consequence of rules alone—conventions which stipulate the meaning of expressions. Empirical events play no role. Only linguistic forms are of interest, not any kind of extra-linguistic fact. Moreover, we have dis-

23 David Cole, personal communication.

24 It might occur to some that dictionaries don't only contain definitions properly so-called (or, as I would have it, hypotheses as to the semantical rules that govern language), but also "general information." For instance, the entry for 'panda' may read 'a large bear-like mammal of the mountains of China and Tibet, with distinctively marked, black-and-white fur'. But, of course, were we to take such a bear out of Asia we would continue calling it 'panda'. So it turns out that the above entry for 'panda' is mostly extraneous information, not at all a report of the strict linguistic rules that govern our use of 'panda'. (Maybe a more apt definition would be a scientific classification, such as *Ailuropoda melanoleuca*.) Such extraneous information pervades dictionaries. Such extraneous information is not *essential* to a dictionary, so it does not count against my thesis that dictionaries are *essentially* reports of semantical rules.

covered inconsistency in Quine's argument on interchangeability and confusion in his description of definition. Quine cannot argue that "necessity statements" amount to "analytic statements" without invoking some notion of synonymy, which he maintains is unclear. Dictionaries, rather than mere reports of observed synonymy, are valuable reports of the semantical rules that govern language. This essay is by no means the final word on analyticity and surely more is necessary before a truly coherent account can be fashioned. We may find solace, however, in the knowledge that such an account is plausible and that we have succeeded in demonstrating the various problems implicit in W. V. Quine's "Two Dogmas of Empiricism".

References

Carnap, Rudolf. *Der logische Aufbau der Welt*. Berlin: 1928.

Carnap, Rudolf. "Meaning and Synonymy in Natural Language." In *Meaning and Necessity*, Second Edition. Chicago: University of Chicago Press, 1956.

Cole, David. "Analyticity" and "Analytic A Posteriori Truth" and personal communications between 9/04 and 01/05. "Analytic A Posteriori Truth" was delivered at the Minnesota Philosophical Society in 2003.

Creath, Richard. "Every Dogma Has Its Day." *Erkenntnis* 35 (1991): 347-89.

Dilman, Ilham. *Quine on Ontology, Necessity, and Experience: A Philosophical Critique*. Albany, NY: State Univeristy of New York Press, 1984.

Ebbs, Gary. *Rule-Following and Realism*. Cambridge, Massachusetts: MIT Press, 1981.

Frege, Göttlob. *The Foundations of Arithmetic*. Evanston, Illinois: Northwestern University Press, 1980. Originally published as *Die Grundlagen der Arithmetik* (1884).

Gibson, Roger F. *The Philosophy of W. V. Quine: An Expository Essay*. Tampa, Florida: University Presses of Florida, 1982.

Gochet, Paul. *Ascent to Truth: A Critical Examination of Quine's Philosophy*. München Wien: Philosophia Verlag, 1986.

Grice, H. P. and P. F. Strawson. "In Defense of a Dogma." *Philosophical Review* 65 (1956): 141-58.

Hookway, Christopher. *Quine*. Stanford, California: Stanford University Press, 1988.

Katz, J. *Semantic Theory*. New York: Harper and Row, 1972.

Kripke, Saul. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1988.

Putnam, Hilary. "The Meaning of 'Meaning'." In *Language, Mind, and Knowledge*, edited by Keith Gunderson. Minneapolis: University of Minnesota Press, 1975.

Quine, W. V. "Carnap and Logical Truth." In *The Ways of Paradox*. Cambridge, MA: Harvard University Press, 1966.

Quine, W. V. "Truth by Convention." In *The Ways of Paradox*. Cambridge, MA: Harvard University Press, 1966.

Quine, W. V. "Two Dogmas of Empiricism." In *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 1953.

THE NO NO-MIRACLES-ARGUMENT ARGUMENT

DANIEL SINGER

University of Pennsylvania

Abstract

The No Miracles Argument is commonly used as a defense of scientific realism. I claim that the No Miracles Argument is begging the question because of the way it uses the notion of "best explanation." I show this by giving a fundamental account of explanation, describing how these explanations can be compared, and showing that, in the case of the No Miracles Argument, the use of the notion of "best explanation" will entail a correspondence theory of truth. I also show that the first premise of the No Miracles Argument and a correspondence theory of truth entail realism. Hence, the No Miracles Argument is begging the question.

Introduction

The No Miracles Argument claims that because any other theory would make the successes of science a miracle, realism must be true.[1] In the following discussion, I will show how the No Miracles Argument uses the concept of explanation. I will give a fundamental account of explanation, which I will show to be common to the major contemporary theories of explanation. I will give an account of how to compare explanations using my fundamental notion. I will then show that the use of "best explanation" in the No Miracles Argument entails that it is begging the question when used as a defense of realism.

"Explanation" and the No Miracles Argument

Recently, one of the most often cited arguments in favor of scientific realism is the No Miracles Argument (NMA). I will take the following to be a canonical version of the NMA as formulated by Matheson:[2]

NMA1) Science has progressed.[3]

NMA2) Scientific realism provides us with a better explanation for this progress than any other philosophy of science.

NMA3) All other things being equal, we should believe the philosophy of science that best explains facts about scientific practice.

NMA4) Therefore, we should believe that scientific realism is true.[4]

---

1 See J. J. C. Smart, *Philosophy and Scientific Realism* (London: RKP, 1963); Richard N. Boyd, "On the Current Status of Scientific Realism?", in ed. Jarrett Lepin, *Scientific Realism* (Berkeley, CA: University of California Press, 1984), 60; Ian Hacking, *Representing and Intervening* (Cambridge University Press, 1983); Robert Nola, "Realism through Manipulation, and by Hypothesis," in ed. S. Clarke and T. D. Lyons, *Recent Themes in the Philosophy of Science* (Boston: Springer, 2002); Carl Matheson, "Why the No Miracles Argument Fails," in *International Studies in the Philosophy of Science* 12, no. 3 (1998).

2 Matheson, "Why the No Miracles Argument Fails," 263.

3 This is generally understood as saying that science has advanced the sum of knowledge we have regarding the things around us and its ability to predict the phenomena that we witness. This is highly contested, but a discussion of this topic is beyond the scope of this article.

4 Other formulations of the NMA may take the conclusion to be "Therefore, scientific realism is true." The

We notice that in NMA3,4 there is an appeal to "explanation" that is not explicated anywhere in the arguments. In this and the next section, I will give an account of what is meant by "explanation" and what we are seeking when we seek the "best explanation." In order to avoid a huge discussion of the philosophy of explanation, I will attempt to provide an account of explanation that is general enough to serve as a foundation of the contemporary philosophies of explanation with the hopes that it will be amenable to all.

At the heart of it, explanation is the answering of a why-question. For instance, the answer to the question "Why does the dog lick itself?" would be an explanation of the phenomenon of the dog licking itself. Now in order to determine what an explanation is, in the very abstract sense, we must happen upon the elements that are essential to answer a why-question. Any answer to a why-question must link the phenomenon in question with other knowledge. For example, an answer to the question of why the dog licks itself might connect our knowledge of the phenomenon with our knowledge of what it means to be a dog. Equally, an explanation of why Socrates was mortal may connect our knowledge of "Socrates is a man" and our knowledge of "All men are mortal." For the purpose of this article, I will use the words "prior knowledge" or "prior belief" to describe a piece of knowledge to which an explanation connects. We can take it as an essential property of any generic explanation E of phenomenon X that E draws a connection between X and a prior belief Q. It is a necessary condition on the notion of "explanation" then that it contains a set of these connections.

It's rather transparent how this model of explanation can serve as a foundation for the many contemporary models of explanation. In fact, each of the major positions simply adds criteria onto the model of explanation I have given: Firstly, the similarities between my model and Hempel's original DN and IS models[5] of explanation are clear as each of those characterizes an explanation as an argument from prior knowledge to the explanandum. The difference is that Hempel extends my model by stipulating that the prior knowledge "be required for the derivation of the explanandum," "must have empirical content," and must be true.[6] Likewise, Wesley Salmon's Causal Mechanical theory of explanation[7] (often cited as the realist response to Hempel's model) adds criteria to my model by stating that the prior knowledge must entail the explanandum (like Hempel's model) but also that the prior knowledge must show how the explanandum "fit[s] into a causal nexus."[8]

While the formulation of Salmon's model is difficult to pin down, it is clear that, as in Hempel's case, Salmon's model is my fundamental model with conditions added to the prior knowledge. It is easy to see how van Fraassen's "constructive empiricism"[9] bases its model of explanation on my fundamental notion: Invan Fraassen's view, an explanation is an answer to a why-question that differentiates the phenomenon from the possible alternatives. For instance, the explanation of why the robber robs banks would tell us why the robber robs banks instead of circuses or tell us why the robber robs at all instead of not robbing. The piece of information that differentiates the explanandum from the alternatives must be prior knowledge such as "the robber enjoys picking vaults." Hence, explanation for

difference between these formulations will be insignificant to the outcome of my argument. It will be clear in the conclusion of my argument how it can be applied to this alternate version.

5 See Carl G. Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," in *Philosophy of Science* 15 (1948), 135-175.

6 Ibid., 153.

7 See Wesley Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton University Press, 1984).

8 Ibid., 9.

9 See Bas C. van Fraassen, *The Scientific Image* (Oxford: Clarendon Press, 1980).

constructive empiricist is an expansion of my notion. Finally, consider Kitcher's unificationist model of explanation,[10] in which Kitcher argues that we count something as an explanation if the prior knowledge used in the explanation adheres to one of the predefined patterns in our current set of explanatory practices. This theory can be viewed as a modification of my fundamental theory, like the others, because it simply limits the set of prior knowledge that can be used in an explanation to those bits of prior knowledge that adhere to one of the established explanatory practices. We see then that the current theories of explanation entail that my fundamental theory of explanation is a necessary (but not sufficient) criterion for explanation. For the sake of this article, I will use the word "explanation" to refer to a set of connections between beliefs or knowledge and the phenomenon to be explained. In the next sections of this paper, I will show that only using this necessary criterion for explanation the NMA can be shown to be begging the question.

Comparing Explanations

In this section, I will expound the idea of "better explanation" by considering how explanations, of the type described above, can be compared. Since it is my goal to show that the NMA's use of "better explanation" entails realism, I must show that the only possible methods of comparing explanations always yield that realism is the best explanation of NMA1. In this section, I will show that there are two ways to compare two abstract explanations, and show that the "qualitative" method of comparison is arbitrary and subjective.

Using the simple conception of an explanation as a set of connections between prior knowledge and the explanandum, we see that one property that differentiates explanations is the cardinality of its set of connections. From this we derive the first method of comparing explanations: quantitative comparison. The explanation above about why the dog licks itself is an explanation that draws only one connection; let's call this an atomic explanation. By combining multiple atomic explanations into a single explanation, we create what I will call a compound explanation. Naturally, we would say, all other things being equal, that an explanation that posits more connections is a quantitatively better explanation since an explanation with more connections would naturally tell us more about the answer to the why-question. I will show in the next section that this method of comparing abstract explanations is irrelevant to the discussion at hand, so I will put this method aside for a moment.

Since comparisons based on the number of connections are dealt with by the prior method of comparison, the second method of comparison must be number-independent. Hence, it must be a property of single connections within explanations. I will call this difference between connections the "qualitative difference." The qualitative difference between two atomic explanations is the difference in the abilities of the explanations to effectively explain the explanandum. For example, for the question of the dog licking himself, there could be two different atomic explanations: one explanation of the phenomenon could say that the dog has an itch and he relieves that itch by licking; the other explanation might posit that the reason why the dog licks himself is that God intends for him to do so. A non-religious person would judge the first to be a better explanation than the second. On the other hand, a person who believes in the ultimate will of the Lord might contend that the second is the better explanation. How do we distinguish which

10 See Philip Kitcher, "Explanatory Unification and the Causal Structure of the World," in *Scientific Explanation*, Philip Kitcher and Wesley Salmon (Minneapolis: University of Minnesota Press, 1989), 410-505.

person is correct? The question comes down to a difference in what I will call meta-explicative values, which I will take to be the values we hold with respect to what makes an explanation qualitatively better than another.

Before I go on to show that one's choice of meta-explicative values (MEV) is arbitrary, I would like to draw a distinction between MEV and the restrictions put on the set of prior knowledge by a formal theory of explanation such as those mentioned above. One may think that, using my vocabulary, the distinction between explanation ρ, an explanation that meets Hempel's requirements, and explanation σ, one that meets Salmon's standards, is a difference in MEV. This is not what I mean by MEV. The qualitative method of comparing explanations is a method meant to describe two particular explanations (rather than abstract explanations in the sense I have been using them). MEV are what allows an actor to distinguish between ρ1 and ρ2, two of Hempel's explanations, or σ1 and σ2, two of Salmon's explanations. The role of distinguishing between ρ and σ is the ongoing discussion of the philosophy of explanation. The point here is that philosophies of explanation distinguish between what is and what is not an explanation, whereas my qualitative method distinguishes only between two bona fide explanations.

Now I will show that the choice of MEV is subjectively arbitrary: Given a formal theory of explanation (such as Hempel's or Salmon's models), let Σ be the set of all explanations of phenomenon χ that satisfy the theory. Using my definition of MEV, we see that a MEV is an ordering principle on Σ, that is, the MEV orders the explanation with respect to which ones are better than others. Now, let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be the set of ordering relations on Σ. Now assume that the choice of ai is not arbitrary (i.e. the choice of MEV is not arbitrary). Then there must exist an ordering relation A that orders the ai with respect to their ability to properly order Σ. To see that this becomes a problem of infinite regress, consider the set $\{A_1, A_2, \ldots, A_m\}$ such that A_i orders $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Notice that as long as Σ has cardinality greater than one, there will be no way of ordering Σ without choosing an arbitrary ordering principle. On the practical level, communities (such as scientists and mathematicians) solve this problem by choosing a reasonable stopping point, but there cannot be anything inherent in the stopping point itself (i.e. actor-independent) that makes the community choose such a point. Hence, I will describe MEV as subjectively arbitrary.

Consider the following objection: When we are deciding who won a footrace, we order the runners by the amount of time it took them to run the race. There is nothing arbitrary about this decision. Hence, by analogy, there is an objectively correct way to order the explanations to determine the best explanation.

The objector is right in saying that this is the correct way to determine the fastest runner. But the reason why this assertion is true is that our community agrees that "being the winner of the footrace" is equivalent to "running the race in the shortest amount of time." The later can be determined by an ordering relation as described by the objector. On the other hand, there is no universal community consensus on what ordering relation is entailed by "being the best explanation," for if there were, the subject of this discussion would already be determined. While it may be the case the some communities may have values that entail a specific ordering relation, I take it to be clear that this is not true across all communities.

Another objection may be inspired by the theory of explanation of van Fraassen. Van Fraassen argues that explanation is a ternary predicate as it involves prior knowledge, the explanandum and the context of the explanation.[11] For example,

11 Van Fraassen, *The Scientific Image*.

two explanations of why the man died may be (1) that he experienced blunt trauma to the head, and (2) that he was the subject of the negligence of his driver. The first explanation is better than the second in a medical setting and the second is better than the first in the courtroom. Then, one non-arbitrary way to order a set of explanations is by how well they explain given a context.

The objector in this case has confused the role of a theory of explanation with the role of MEV. It is not the case that (1) is better than (2) in the medical setting; rather, in the medical setting (1) is explanatory while (2) is not. It is the role of the theory of explanation, not MEV, to determine whether a set of connections is or is not an explanation.

I have now shown that given my fundamental framework of explanation, there are two ways to compare explanations, quantitatively and qualitatively, and that qualitative comparisons of explanations rely on a choice of a subjectively arbitrary set of meta-explicative values. In the following sections, I intend to show that a correspondence theory of truth is entailed by the NMA's use of "best explanation," and through an explication of the concept of realism, it will be clear that a correspondence theory of truth along with NMA1 entails realism, which leaves the NMA begging the question.

Realism, Correspondence and the NMA

In this section, I will show that the acceptance of the first premise of the NMA and a correspondence theory of truth entail the acceptance of scientific realism. This is important to my argument since in the following sections, I will argue that the NMA's notion of "better explanation" will always be one with a correspondence theory of truth. I will also show that the NMA's "better explanation" must be one that is a qualitatively better explanation, rather than a quantitatively better explanation.

So, what is scientific realism? While scientific realism can take many forms, there are key tenets without which no picture would be realist. As Putnam puts it, the realist holds that "the world consists of some fixed totality of mind-independent objects. There is exactly one true and complete description of 'the way the world is.' Truth involves some sort of correspondence relation between words or thought-signs and sets of things."[12] I will take this to be the staple realist contention. Kirkham characterizes a correspondence theory of truth as an isomorphic mapping between the truth and the facts of the (mind-independent) world.[13] The correspondence theorist holds that sentences such as "the dog licks itself" are truthful in so much as the concepts contained within them map onto the things, properties, relations, etc. that compose the world. If the correspondence-truth theorist holds that P is true, then he believes that there is an isomorphic map between the objects described by P and objects in the world.

Consider the actor who accepts NMA1 and a correspondence theory of truth. It would be difficult to exactly delimit the claim made in NMA1, but we can at least accept that NMA1 grants that either (1) the theories we have now are closer to the truth than their predecessors or that (2) they make better predictions than their predecessors. Kitcher would go as far as to say that this means that the contemporary theories are approximately true.[14] If we take it that NMA1 says that (1) the

12 Hilary Putnam, *Reason, Truth and History* (Cambridge University Press, 1982), 49.
13 Richard L. Kirkham, *Theories of Truth: A Critical Introduction* (Cambridge, Mass.: MIT Press, 1992), ch. 4.
14 See Philip Kitcher, "On the Explanatory Role of Correspondence Truth," *Philosophy and Phenomenological Research 66* (2002), 357-359.

theories we have now are closer to the truth, then it suffices to consider that since every scientific theory posits or makes mention of some mind-independent object, it is fair to assume that the person who accepts NMA1 accepts that a true scientific theory, in so much as one could exist, posits relations between mind-independent objects. Then, since the actor accepts a correspondence theory of truth, he must concede that these objects posited by the true theory actually exist in the world, which makes the actor a realist. On the other hand, if we take it that NMA1 claims that (2) new scientific theories make better predictions than their predeccesors, we run into the following situation: the actor accepts that some things that the scientific theory says are true, namely the predictions, but presumably, the actor witholds judgement on the remaining claims of the theory. But, the claims of a scientific theory, I take it, are the statements that are true of the class of models picked out by the theory (on the semantic view of scientific theories). Then, on this account, the actor must demarcate which of the sentences that are true of the scientific theory are predictions and which are not. To provide such an account would essentially amount to reviving the received view of theories on which observable sentences are distinguishable from theoretical sentences. On its face, it appears that this is not an attainable goal. So, the actor who wishes to claim that NMA1 means that newere scientific theories give us better predictions must either accept as true all of the sentences that are true of the scientific theory (which would put him in the campe of naive realists) or reject the scientific theory, which would not allow him to even claim that some predictions of theory are true. It appears then that if an actor accepts NMA1 and a correspondence theory of truth, she would accept realism.[15] [16]

Now it simply remains for me to show that the notion of explanation used by the NMA entails a correspondence theory of truth. Notice that the concept of "better explanation" as referenced in the NMA is actually an appeal to a *qualitatively* "better explanation," rather than quantitatively better or a combination of both. This is true by the following proof: By the definitions above, explanation X is better than explanation Y only if X is quantitatively better than Y, X is qualitatively better than Y, or both. If it were not the case that "better explanation" means "purely qualitatively better explanation," then we must either conclude that it refers to a quantitatively better explanation or a both quantitatively and qualitatively better explanation. Let the *explanation class* of theory T be the class of all explanations employed by T to explain the scope of T. Using this tool, we can see that "better explanation" as used by the NMA cannot mean purely quantitatively better: Let α be the cardinality of the explanation class of scientific realism and β be a larger cardinality. Notice that we can always create an explanation class of a (non-realist) theory such that the cardinality of the class is β. For example, we could create the theory GL that posits β many Gods and for every phenomenon that NMA claims that realism explains, GL explains it by saying that each of the β many Gods intend that phenomenon. If we assume there are a countable number of phenomena, then the cardinality of the explanation class of GL is β. If "better explanation" were meant in a purely quantitative sense, the supporter of the NMA

15 For a more detailed (and stronger) argument of this sort see Kitcher, "On the Explanatory Role of Correspondence Truth."

16 Two notable exceptions to this exist. McTaggart makes an argument in which he proclaims to be able to be a realist without a correspondence theory of truth. See John McTaggart, *The Nature of Existence* (Cambridge Univesity Press, 1921). Secondly, the internal realist (see Hilary Putnam, *Reason, Truth and History*) would hold NMA1 without assenting to the assertion that NMA1 entails the existence of mind-independent objects. For this purposes of this discussion, I will ignore these positions.

would have to conclude that GL can provide a better explanation of NMA1 than realism, but clearly the arguer would reject that. This shows that the quantitative method of comparing explanations is essentially trivial in the case of the NMA. Consequently, we know that "better explanation" as understood in the context of the NMA refers to qualitative difference.

Since we now know that the notion of "better explanation" as used by the NMA is that of a qualitatively better explanation, we turn our attention to the set of meta-explicative values that define the NMA's qualitative notion of "better explanation." In the following sections, I will attempt to determine what set of meta-explicative values, V, is assumed by the NMA to determine that realism is the best explanation of NMA1. I will then show how this assumption of an assignment of V always entails a correspondence theory of truth.

The MEV assumed by the NMA

There are two ways we can figure out what V, the set of meta-explicative values used in the NMA, is: we can (1) deduce from the premises of the argument what V must be such that it makes the other premises and conclusion true, or (2) we can posit reasonable options for V based on the way the argument is used by contemporary philosophy and science. Obviously the first method is preferable as its results are deductively valid, but unfortunately it will not work as shown by the following argument: Assume there is a v_1 such that v_1 picks the realist explanation over any other explanation for the success of science and v_1 picks out W as the second best explanation. Now there can also be a v_2 such that v_2 picks out realism as the best explanation for the success of science, but v_2 always assigns explanation W as the worst explanation. Clearly, $v_1 \neq v_2$, but v_1 and v_2 would both make the other premises of the NMA true. Therefore, deduction to correct specific assignment of V is not possible.

One may propose that deduction to an equivalence class of possible assignments of V, rather than a single assignment of V, is possible, and this class, for the purposes of this discussion, can be treated as an assignment of V. In other words, one can deduce the set V′ of all assignments of V such that $\forall v' \in V'$, v′ treats realism as the best currently existing explanation for the success of science.

The proposition is right that this suggestion is a possible solution to the question of how to delimit V, but his assumption that this strategy will not affect the discussion is false by the following argument: Assume that there is a v_r such that v_r is a meta-explicative value that entails that realism is the best possible explanation for anything. This possible assignment of V, if it is the one being employed by the arguer of the NMA (ANMA), is uniquely significant to this discussion in that it makes it clear that the ANMA is begging the question by using the NMA. On the other hand, there could be a $v_{\sim r}$ such that $v_{\sim r}$ gives an equal opportunity to all philosophies of science that attempt to explain the success of science, but $v_{\sim r}$ entails that realism is the best explanation of this success because it has the shortest name when written in Sanskrit. v_r and $v_{\sim r}$ would be in an equivalence class of the type defined by the objector, but clearly v_r begs the question of the NMA and $v_{\sim r}$, even though it's a strange value to hold, does not. So, treating an equivalence class of values as V is not sufficient for this discussion.

For determining V, we are now left with the second option, which determines V by appeal to the way contemporary philosophers of science intend for "better explanation" to be interpreted. Seemingly, there are two possible choices for V by the standard usage of explanation: (1) simplicity and clarity, and (2) a subjective measure of the likelihood the truth of an explanation, something I will call "truth probability."[17] By simplicity (1), I mean the ability of an explanation to delimit which pieces of information are being connected and clearly define the nature of the connections. By truth probability (2), I mean a subjective valuation proportional to the probability that the explanation could be true given the evaluator's previous knowledge base. This second criterion can be conceived of as comparative feasibility. I will take for granted that when philosophers judge the quality of an explanation they judge it with respect to one or both of those options.[18]

I have argued that the NMA's notion of "better explanation" must be one that uses a qualitative comparison, and I have provided two options for the MEV in use by the qualitative comparison. In the next section, I will argue that both of these MEVs, when used to compare explanations, will always prefer an explanation that employs a correspondence truth. Since I have argued above that correspondence truth together with the first premise of the NMA entail realism, it will be shown that the NMA begs the question.

How V Entails Correspondence Truth

In the previous section, I argued that the ANMA must be using either simplicity or truth probability as the MEV for comparing explanations. I will now show that both of these options for V leave the ANMA begging the question since he is using a denition of explanation that entails that realism is the best-suited philosophy of science under any circumstances: (1) Suppose that the ANMA intends the first possible assignment of V (simplicity and clarity), that "best explanation" is meant in the sense of simplest and clearest explanation. I contend that an explanation that uses a correspondence truth is the simplest and clearest possible explanation of NMA1. NMA1 says that science has progressed and has been successful. In doing so, science has produced theories that posit structures that act, in the scientific model, as the cause of the progression of science.[19] So what is the simplest possible explanation of this success? Surely, the simplest explanation of why the models are predictively accurate is that the things and relations posited in the model directly map onto the things and relations in reality (I will call this the "identity map"). This must be the case; consider this argument by reductio ad absurdum: Assume there is a theory that is both simpler than the identity mapping and does not include the identity mapping (because doing so would make it the identity mapping or obviously more complicated than the identity mapping). This theory would have to say that at least one of the things posited by the models does not directly map onto reality or does not exist in reality (because if it didn't say such a thing, it would be the identity mapping). If the theory does not map everything to its real counterpart, then the theory must explain why that element or relation exists in the model but does not exist in the world, which would entail either that it is not an explanation (in the case that it does not draw a connection to explain that part of the model) or it is more complicated than the identity mapping. This is a contradiction with the assumption; hence, the simplest explanation of the success of science is the identity mapping, which is an application of a correspondence theory of truth.

17 Notice the similarity of these notions and the commonly accepted Ockham's Razor.

18 I am not currently able to produce an argument that these are the only two standard assignments of V, but it seems clear to me that any choice of meta-explicative value here will be subject to similar arguments to those below.

19 Here I am employing the standard semantic conception of scientific modeling, which identifies scientific models with logicians' models. The structure then of scientific models is identified with the relations between the elements of the model, which are often used to represent causal forces in the model.

If it is the case that the ANMA intends to appeal to the second possible assignment of V, truth probability, then we will see that the best possible option for an explanation will also be one that employs a correspondence theory of truth: Using the same method used for the first option, let us consider the best possible explanation that could satisfy this criterion. In order to do so, we must decide what would be the best criteria for determining how likely something is to be true. Surely, when comparing one explanation that posits something that is commonly witnessed or accepted with an explanation that posits something that rarely if every witnessed or accepted, we think that the former is better because of its higher likelihood to be true. In other words, if one explanation posits something that we perceive as commonplace, we believe it is more likely to be true than an explanation that posits something rare. We see then that our valuation of the truth probability of an explanation is proportional to the frequency with which we believe the things posited by the explanation are experienced. Now notice this peculiar property of experience: We say that we experience X when X is presented to us. After having an experience of X, the object X *corresponds* to our experience of X. We see that when we have any generic experience, we have experience of correspondence. Therefore, nothing is more experienced than correspondence, and hence, correspondence has the maximum value of truth probability. Likewise, an explanation that assumes correspondence between the entities of the phenomenon and the entities of the world would be qualitatively the best possible explanation when V is a measure of truth probability or feasibility.

Conclusion

Under either potential value that could be appealed to by the ANMA, the best possible explanation is one that appeals to a correspondence theory of truth. In addition, as we saw above, if an actor using a correspondence theory of truth accepts NMA1, then he is a realist; hence, we see that the ANMA is begging the question in using the NMA. Admittedly, my argument has a weakness in Section 4 because it would not be feasible to account for all possible assignments of V, but I leave it the ANMA to show that there is some commonly accepted MEV that makes realism the best explanation of NMA1 without entailing a correspondence theory of truth.

The ANMA may claim that I have not shown that he is begging the question; I have only shown that his argument is valid: given that the premises are true, I have shown that the conclusion is true. However, in the fight over realism, a satisfactory argument would not be one that is trivially true. I have shown conversely that the NMA is true syntactically, and hence it is not a satisfactory argument for the defense of realism.

I have shown that "better explanation" as used by the ANMA is coextensional with "closer to scientific realism." It is now clear that the ANMA is begging the question if we use substitution of these concepts in the canonical of the NMA given above:

NMA*1) Science has progressed.

NMA*2) *Scientific realism provides us with a theory closer to scientific re alism than any other philosophy of science.*

NMA*3) *All other things being equal, we should believe the philosophy of science that is closest to scientific realism,* NMA*4) Therefore, we should believe that scientific realism is true.

Under this substitution, the NMA is a silly circular line of reasoning. It would take a miracle to conclude that this argument is not begging question.

Finally, I would like to address what I will call the naive NMA. The naive NMA claims that because any other theory would make the successes of science a miracle (and being a miracle is not an explanation), scientific realism is true. The naive NMA is different from Matheson's NMA because the naive argument implies that *the only explanation* of NMA1 is a realist theory. Notice that NMA2 implies that there are more than one explanations of NMA1, only one of which is scientific realism. The argument provided above cannot address the naive NMA because there only one way to order the one possible explanation posited in the naive argument, so my assertion that the choice of MEV is arbitrary no longer holds. By arguing the naive NMA, the discussion of the NMA has been moved to the realm of philosophy of explanation. The theory of explanation held by the arguer of the naive NMA would have to say that the only explanation of NMA1 is realism and that nothing else can serve as an explanation of NMA1, which is opposed to the NMA's implication that other theories explain NMA1, but not as well. This fact would have to be a byproduct of the arguer's theory of explanation, but it seems to me that most people who make this argument are forming their theory of explanation to guarantee this result without otherwise justifying it. This is the philosophical equivalent of sticking your fingers in your ears and screaming "No, No, No;" though, if the arguer of the naive NMA could provide a sound theory of explanation that entails this result, my argument would not be able to show that argument to be circular.

References

Boyd, Richard N. "On the Current Status of Scientific Realism?" In *Scientific Realism*, edited by Jarrett Lepin. Berkeley, CA: University of California Press, 1984.

Hacking, Ian. *Representing and Intervening*. Cambridge University Press, 1983.

Hempel, Carl G. and Paul Oppenheim. "Studies in the Logic of Explanation." *Philosophy of Science 15* (1948), 135-175.

Kirkham, Richard L. *Theories of Truth: A Critical Introduction.* Cambridge, Mass.: MIT Press, 1992.

Kitcher, Philip. "Explanatory Unification and the Causal Structure of the World." In *Scientific Explanation*, Philip Kitcher and Wesley Salmon. Minneapolis: University of Minnesota Press, 1989, 410-505.

Kitcher, Philip. "On the Explanatory Role of Correspondence Truth." *Philosophy and Phenomenological Research 66* (2002), 357-359.

Matheson, Carl. "Why the No Miracles Argument Fails." *International Studies in the Philosophy of Science 12*, no. 3, 1998.

McTaggart, John. *The Nature of Existence*. Cambridge Univesity Press, 1921.

Nola, Robert. "Realism through Manipulation, and by Hypothesis." In *Recent Themes in the Philosophy of Science*, edited by S. Clarke and T. D. Lyons. Boston: Springer, 2002.

Putnam, Hilary. *Reason, Truth and History*. Cambridge University Press, 1982.

Salmon, Wesley. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, 1984.

Smart, J. J. C. *Philosophy and Scientific Realism*. London: RKP, 1963.

van Fraassen, Bas C. *The Scientific Image*. Oxford: Clarendon Press, 1980.

THE HEART'S REASONS: INTUITION AS AN AUTHORITY IN PRACTICAL REASONING

RONNI SADOVSKI
Swarthmore

Since the seventeenth century, Western philosophy has tended to polarize, not only mind and matter but also reason and feeling — to treat these as separate aspects of life, not intelligibly related. — Mary Midgley[1]
The heart has reasons that reason cannot know. — Blaise Pascal[2]

The dichotomization of mental life into opposed dominions of reason and feeling is a staple of our cultural inheritance. But although we speak of feelings as if they were irrational, we recognize that they often serve as the reasons for our actions. Whenever someone investigates a hunch or goes with her gut, she is letting her intuitive feelings guide her decision. Unlike desires or tastes, intuitions are not treated merely as idiosyncratic preferences for one action or another; rather, we treat the heart (and the gut) as a compass, responsive to facts about the world in a way that the mind often cannot be. Intuitive feelings are not merely *related* to practical reasoning; in fact, they are intrinsic to it. My goal is to make the relation between intuitive feeling and practical reasoning intelligible.

I will argue that intuition plays an indispensable role in proper reasoning — the role of an authority. Like an expert's testimony, intuitive feelings serve as authoritative reasons to judge and act in a certain way. And like every authority, intuition has its limits. We advise our friends to listen to their hearts just as often as we instruct them to use their heads. Through my account of intuition's authoritative role, I hope to explain why we are justified in heeding our feelings, and also to mark out the limits of their authoritative legitimacy.

The Structure, Justification and Scope of Authority

In giving this account, I will rely on the theory of authority given by Joseph Raz in his book, *The Morality of Freedom*. Raz carves legitimate authority into two constitutive properties: preemption and dependence. The directive of an authority is *preemptive* when "the fact that an authority requires performance of an action is a reason for its performance that… should exclude and take the place of some of the [other reasons relevant to the decision]".[3] For instance, if my mother asks me to be home by 7:00, she surely will have done so for some reason. But if I come home on time because I find that reason compelling, I will merely be complying with her request, and not obeying it. My mother's directive is only authoritative if I come home on time *because she told me so*. When a directive motivates our action authoritatively, it preempts the reasons that motivated the authority to issue the directive in the first place.

How could the bare fact of someone's being an authority justify our obedience to him? One central part of that justification is the second constitutive property of an authoritative directive: *dependence*. A directive is dependent when it is "based

1 Mary Midgley. *The Ethical Primate: Humans, freedom and morality*, (New York: Routledge, 1994), 13.
2 Robert C. Solomon, *The Passions: Emotions and the Meaning of Life*, (Indianapolis: Hackett, 1993), 58.
3 Joseph Raz, *The Morality of Freedom*, (New York: Oxford, 1986), 46.

on reasons which already (1) independently apply to the subjects of the [directive] and (2) are relevant to their action in the circumstances covered by the directive".[4] My mother's request that I be home before seven is dependent if, say, my parents plan to go out at seven and I am responsible for taking care of my little sister while they're gone. My responsibility as a member of the household is a reason that applies to me independently of my mother's request, so the curfew satisfies the first component of Raz's criterion of dependence. The responsibility is relevant to the action that she asked me to perform because I cannot meet my responsibility if I don't come home on time. Thus, the curfew satisfies the second component of the dependence criterion. Because it meets both the criterion of preemption and that of dependence, this curfew constitutes a legitimate exercise of my mother's authority.

All authorities are fallible, and even the best of them will occasionally make mistakes in evaluating the reasons that apply to their subjects. Raz permits that an authority can fail to meet the criterion of dependence some of the time and still be legitimate, as long as it acts on dependent reasons often enough.[5] But this caveat highlights an apparent problem arising from the conjunction of authority's preemptive and dependent properties: if an authoritative decree depends on reasons that already apply to its subjects, why should the decree preempt those reasons? Why shouldn't the subjects simply act according to those reasons that apply to them and ignore the authority altogether? In other words, what justifies the authority's authoritativeness?

Raz's answer to this question forms the crux of his explication of authority's legitimacy: the Normal Justification Thesis. A would-be authority (whose decrees are preemptive and dependent) legitimately commands its would-be subject when

> the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which apply to him directly.[6]

If my mother's authoritative decrees are dependent on reasons having to do with, say, my household responsibilities, then they are authoritative as long as obeying them improves my ability to respond to those reasons. In other words, her authority is legitimate if and only if I meet more of my household responsibilities when I let her rules guide my actions than I would just by trying to identify and meet my responsibilities by myself.

In one breath, the Normal Justification Thesis legitimizes authority and restricts its scope. Where Normal Justification obtains, the authority serves the interest we have in responding properly to the reasons that apply to us, and (barring extenuating circumstances) we are clearly right to obey its commands. Where Normal Justification does not obtain, however, the authority does nothing for us and cannot legitimately command us.

Challenging Authority

Here, Raz's explanation of authority raises another apparent paradox. How can a subject determine whether the authority that commands her meets the criterion

4 *Ibid*, numerals mine.
5 *Ibid*, 47.
6 *Ibid*, 53.

of Normal Justification? Her only option, it seems, is to use her own best reasoning to evaluate the authority's decree. But if she obeys the authority only when it coheres with her own independent reasoning, then the Normal Justification Thesis is empty, and the authority is doing no work at all. The problem grows stickier still when we notice that even an authority that once met the criterion of Normal Justification may suddenly cease to do so. How can a subject ever justify her belief in an authority's legitimacy?

Though it seems intractable, this problem actually poses no major difficulty for Raz. Consider the authority of a pocket calculator. When I rely on my calculator as an authority, I am much more likely to do arithmetic correctly than I would be if I were to work math problems out by hand. Hence, my calculator meets the criterion of Normal Justification. I do not need to check each operation by hand to be justified in this belief, and in fact, if my calculator were to suddenly start malfunctioning, I might accidentally continue to accept its authority after Normal Justification has ceased to obtain. On the other hand, if my calculator told me that the sum of two even numbers is an odd number, or if I suddenly found myself inexplicably getting questions wrong when I used that calculator to do math homework, I would have reason to believe that it no longer meets the criterion of Normal Justification. Such an event would drive me to check its calculations by hand and reevaluate the legitimacy of the calculator's authority.

For Raz, an obvious error on the part of an authority gives its subject a good reason to investigate the authority's performance with respect to the Normal Justification Thesis. There are also a number of other grounds on which an agent might challenge the legitimacy of an authoritative decree. For example, an authority may fail to respond to all the reasons it purports to preempt. If a judge systematically fails to consider divorcees' assets in awarding alimony, the legitimacy of his authority can and should be called into question. A would-be authority can also be challenged for issuing a directive on the basis of reasons that are not appropriately related to the actions it requires of its subjects. *Catch-22*'s General Peckem and Colonel Scheisskopf are guilty of this abuse of authority when they order their fighter pilots to bomb towns in order to create bomb patterns that will look pretty in aerial photographs. An authority could also overextend its command by issuing orders in domains where it doesn't belong. My mother's authority would be subject to this objection if she tried to give my friends curfews. Even if an authoritative decree is legitimate, the reason it gives to act in a certain way may be defeated by a sufficiently strong reason to do otherwise. A professor may require that I write a paper by a certain date, but if some emergency befalls my roommate and she urgently needs my help, her needs may override the professor's (albeit legitimate) demands.

My analysis of intuition's authoritative role in reasoning will justify its legitimacy using the same criteria that Raz uses to justify the legitimacy of authority in general. With these justifications come the same limitations that apply to the authorities Raz describes. Like any authoritative decree, an intuition can be challenged for any of the reasons enumerated above, and should not be taken to guide action outside of the strictly limited scope of its legitimacy.

Practical Reasoning and Intuition

Before I can apply Raz's theory of authority to intuition, I need to say a few words about the structure of practical reasoning. In Raz's account of authority's legitimacy, he makes reference to *reasons that apply to subjects*. I will use the same language — there exist reasons (in the world, as it were) which apply to us and which we ought to acknowledge and obey by performing certain actions. These reasons are different from the reasons that motivate our actions; whereas the *reasons that apply to us* exist independently of our acknowledging them, the *reasons for which we act* are psychological entities, which attain their status as reasons by contributing to our deliberation. These, in turn, are different from the deliberative process itself, which combines and balances all of our psychological reasons to act and produces an overall *judgment* about what to do. For clarity's sake, I will use the term "f-reasons" to refer to the reasons that apply to us (factive reasons), "m-reasons" to refer to the reasons for which we act (motivating reasons), and "judgment" to refer to the output of the deliberative process. Note that, by the account given earlier, authorities depend on our f-reasons and provide us with m-reasons to act.

Where do feelings fit into this picture? Although our inherited wisdom tells us that feelings are opposed to reasons, I will treat feelings instead as a type of reason. Specifically, feelings can be either m-reasons or judgments. For instance, when my brother visited Cornell and Stanford and found that he felt more at home among Stanford students, his feeling was an m-reason to prefer Stanford. This reason responded to certain f-reasons having to do with his personality and various details of Stanford's social culture, and it ultimately became a component of his judgment that Stanford was his best choice. That judgment also came to him as an intuitive feeling: my brother could not have given numerical weights to his various m-reasons to prefer each institution, nor could he have given a conclusive argument that explained why the balance of his m-reasons ultimately sided with Stanford. Nevertheless, when he considered all of the m-reasons to prefer Stanford alongside the m-reasons to prefer Cornell, Stanford felt right. His judgment was a matter of intuition.

Although all intuitive feelings are reasons, it is clear that not all reasons are intuitive feelings. Robert Solomon distinguishes two types of reasons: prereflective and reflective.[7] Reflective reasons are the kinds of reasons philosophers usually deal with. When Socrates interrogates Euthyphro about his decision to prosecute his father, for example, he is asking for an account of Euthyphro's reflective m-reasons: that his father killed a man unjustly, that it is pious to prosecute murderers, and so on. Like m-reasons, judgments can also be reflective. Utilitarianism gives its adherents a prescriptive methodology for reflective judgment: consider all the reasons that have to do with the action's consequences, and weigh those reasons based on the principle of utility maximization. Whatever we may say about utilitarianism, it surely is the very model of reflective judgment!

Prereflective reasons, on the other hand, are intuitive — they are *felt*, rather than thought through. Just like reflective reasons, prereflective reasons may be m-reasons or judgments. When Raskolnikov confesses to murder in Dostoevsky's *Crime and Punishment*, he is acting on a prereflective judgment, informed by prereflective m-reasons. He cannot put his m-reasons into words or articulate his judgment as a logical argument, but his judgment and his m-reasons are nevertheless responsive to the facts of Raskolnikov's situation — his f-reasons.

It is important to note that this difference is not only a phenomenological one, but a difference in the represented content of the reasons. Since reflective m-reasons are articulated and prereflective m-reasons are not, the two types of reasons differ in the way that they refer to f-reasons. We might say that reflective m-reasons *point* to the f-reasons they depend on, whereas prereflective m-reasons

7 Solomon, 182.

can only *gesture*. Contrast a reflective m-reason to avoid a used-car dealer ("he lied about the car's gas mileage") with a prereflective m-reason to do the same ("there's something fishy about him"). Both depend on f-reasons to distrust the man, but the reflective m-reason explicitly refers to the f-reason it depends on, while the prereflective m-reason does not.

The distinction between reflective and prereflective judgment is similar: a reflective judgment is like an argument that explains why the m-reasons that the agent is considering support one action more strongly than any other. This is not to say that the agent who makes a reflective judgment needs to make such an argument in words, or even use words to think it through. If her judgment is to be reflective, the agent must *reflect* on the justification for deriving her conclusion from the m-reasons at hand, but she need not use any particular mental device to do it. Like a reflective judgment, a prereflective judgment follows from some chain of justification. But that justification is not available to the agent who makes a prereflective judgment.

The fact that prereflective reasons do not have explicit referential content might make us wonder why we obey them. A reflective m-reason *tells* us why it should motivate our action — it says that such-and-such an f-reason applies to us. A reflective judgment also wears an explanation on its sleeve. But a prereflective m-reason or judgment can only prod the agent in one direction or another; the only justification it gives is "because I say so." Perhaps it is this mysterious quality of intuition that has made some philosophers so suspicious of its legitimacy. At any rate, intuition's prereflective nature invites us to treat it as an authoritative decree. Having distinguished reflective reasons from prereflective reasons, I am finally equipped to examine intuition using the framework of authority developed earlier.

Intuition's Authority

In Raz's theory, the decree of a legitimate authority is an m-reason that depends upon (and preempts) some set of f-reasons that apply to its subject. Its legitimacy depends primarily on the Normal Justification Thesis, which we can now rephrase using the terminology of reasons established above: an authority is legitimate if and only if its subject is likely better to comply with the f-reasons that apply to him if he takes the authority's directives as m-reasons that motivate his action, instead of trying to discover those f-reasons for himself and thereby formulate m-reasons independently of the authority.

Phrasing the Normal Justification Thesis in terms of this ontology of practical reason should help to clarify the link between authority and intuition. This link is clearest with respect to prereflective m-reasons (though, as we shall see, it also has interesting implications for prereflective judgments). Just as I obey my mother's authority because I will respond better to my preexisting f-reasons by heeding her than I would on my own, I listen to my heart in those situations where I have reason to believe that I will respond to my f-reasons better by doing so than I would have if I had analyzed my circumstances to pieces. The observation that intuition may be legitimated by the Normal Justification Thesis lends some promise to the enterprise of cloaking intuition in authority's robes, but by itself it is insufficient to defend my thesis. To count as an authority in Raz's sense, intuition must be dependent and preemptive. In what follows, I will show that dependence and preemption are constitutive properties of intuitions.

The dependence of intuition is what differentiates it from a preference or an impulse. A preference for year-round sunny weather over cooler climates is just an f-reason (which, when recognized by the agent, gives rise to an m-reason) to live in a warm climate. One need not like sunny weather for any particular reason; it can just be a taste one has. Intuitions are unlike mere preferences in that they respond to f-reasons that apply to the subject independently of the intuition's existence. If my brother feels more at home at Stanford than he does at Cornell, his feeling is more than just a preference for Stanford, which he would weigh alongside all of his other m-reasons to go there or to Cornell; rather, it depends on a number of f-reasons independent of the feeling itself — reasons having to do with the sunny California weather, the southwestern architecture of the campus buildings, or the students' relaxed social culture. Intuitive feelings are dependent reasons in that they respond to facts.

Intuitions also preempt those facts that they respond to. Suppose my brother feels apprehensive about applying to Dartmouth. He may realize that his apprehension arises from his belief that Dartmouth's fraternities run the college's social scene. But my brother cannot now consider himself to have two separate reasons to not to apply to Dartmouth: first, the extent of its Greek life, and second, his feeling of apprehension. My brother's opinion of Dartmouth's fraternities can explain his feeling, but it does not motivate his decision *alongside* his apprehension. If he takes his apprehension as a reason against applying to Dartmouth, it must be a preemptive reason. This is not to say that he must take his apprehension as a preemptive, prereflective m-reason. He may ignore it instead, and take Dartmouth's Greek life as a reflective m-reason not to apply there. The point is that he cannot do both. An intuition cannot motivate action alongside the reasons it depends on. It can serve as a preemptive reason, or no reason at all.

Given that choice, when should an agent go with his gut, and when should he think things through? If intuitions really are authoritative, then Raz's Normal Justification Thesis provides the answer to this dilemma: an agent should treat his intuitions as preemptive m-reasons whenever doing so improves his ability to respond to the f-reasons that apply to him. Hence, we should let our intuitions preempt the f-reasons they depend on when they are the best way to respond to those reasons, but we should ignore them if we have good grounds to suspect that reflective reasoning would serve us better. The kinds of grounds one can use to challenge any traditional authority (obvious failure to meet the requirements of Normal Justification, failure to respond to all the reasons it would preempt, and so on) can also undermine the legitimacy of someone's intuitions. And just like any other authority, a prereflective m-reason can be defeated by any sufficiently strong opposing m-reason. Even if my brother feels certain that Stanford is the place for him, a meager financial aid package might force him to look elsewhere.

The Next Step

I have shown that the relation between a prereflective m-reason and the f-reasons behind it is just like the relation between any authoritative decree and the reasons that it depends on. But not all intuitions are prereflective m-reasons. Some intuitions are prereflective judgments. Can intuitive judgments also be authoritative?

Because Raz formulated his theory with external authorities in mind, his account deals only with authoritative m-reasons that preempt f-reasons. But my analysis deals with authorities internal to a single agent's reasoning process, and therefore makes it possible for us to extend the domain of authority from m-reasons to judgments. Prereflective judgments are just as dependent and preemptive as prereflective m-reasons are: they depend upon and preempt the m-reasons that relate to the performance of some action. While judgments cannot, strictly speaking, meet the criterion of Normal Justification (because Raz's Normal Justification Thesis is formulated in terms of f-reasons and m-reasons), they can meet an analogous criterion of justification: a prereflective judgment is legitimately authoritative if and only if the agent will be better able to assess the m-reasons relevant to the action he is deliberating about if he takes his intuition as an authoritative, instead of trying to weigh his m-reasons reflectively. Since this criterion is analogous to the criterion of Normal Justification, it seems to me that, together with the dependent and preemptive properties of intuitive judgment, it gives us sufficient grounds to treat intuitive judgments as authoritative.

This analysis of intuitive judgment shows how Raz's theory of authority might be extended into new domains. My argument that prereflective m-reasons are authoritative showed that one part of a single agent's reasoning can play an authoritative role with respect to another part of his reasoning. Extending this idea to judgment requires us to revise Raz's criterion of Normal Justification, but also enables us to generalize his theory in an interesting way, making it more broadly applicable to the problems of practical reasoning. If my argument about the authoritative quality of intuitive m-reasons and judgments has been at all compelling, its success suggests that a broader treatment of authority's role in practical deliberation may be a productive philosophical project.

References

Midgley, Mary. *The Ethical Primate: Humans, Freedom and Morality*. New York: Routledge, 1994.

Raz, Joseph. *The Morality of Freedom*. New York: Oxford, 1986.

Solomon, Robert C. *The Passions: Emotions and the Meaning of Life*. Indianapolis: Hackett, 1993.

THE SEMANTIC VIEW AND THE *α*-MODEL

RICHARD LAWRENCE
University of Pennsylvania

Abstract

In what follows, I discuss the semantic view of scientific theories, specifically in the 'state space' formulation that has been developed by Bas van Fraassen, Frederick Suppe, and others and applied by Elisabeth Lloyd. I consider the claims that the state space view makes about how scientific theories are best understood. I then discuss a particular model from network theory, the *α*-model developed by Duncan Watts, and try to apply the state space view to the *α*-model theory. I argue that the way Watts uses the *α*-model is not described very easily under the state space approach, because its parameters are idealized in a way that the state space approach does not account for, and because the relation between the *α*-model and the world is not one of statistical fit between its models and data.

Introduction

Some recent discussion in philosophy of science has centered around the issue of scientific *modeling*. Modeling has generally been recognized in this discussion as the practice of creating simplified, idealized, or abstract representations during the course of empirical work. Philosophers have asked what models are, how they are

constructed, whether or not they actually represent empirical phenomena, how they can describe or explain phenomena, and so on.

The major advocates for understanding how models function in scientific theories have been proponents of the *semantic view of theories*, a philosophical position that holds that theories are to be understood directly in terms of the models they present, instead of in terms of their particular formulations in language. This basic

tenet is what makes the view a *semantic* one: it doesn't presume that scientific theories can be expressed in a "language of science," in which all theoretical descriptions of the world are (syntactically) derivable from statements about empirical data, as the logical positivists tended to believe. By holding that scientific theories are independent of their linguistic formulations, proponents of the semantic view hope to escape the complicated issues surrounding syntax, "observational" versus "theoretical" concepts, and rules of inference in science that plagued the logical positivists before them.

The semantic view, as developed by Patrick Suppes, Bas van Fraassen, Frederick Suppe, and others, has evolved into two different approaches to analyzing scientific theories. The first, due largely to Suppes' work, is a set-theoretic approach, where models are viewed as *possible realizations* of theories and data. In Suppes' view, theories and actual data are related via a hierarchy of such models.[1] The

1 Suppes, Patrick. "Models of Data." In *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, Edited by E. Nagel, P. Suppes, A. Tarski. New York: Springer.
Suppes, Patrick. "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences." *Synthese 12*, 287-301.

second approach, developed by van Fraassen, Suppe, and others, views theories as structures that specify certain parameters for describing phenomena. These parameters are arranged as the axes of a *state space* or *phase space*. A model, on this view, is a particular assignment of values to the parameters of the theory; it is therefore seen as a point in, or trajectory through, the state space. Observation data is plotted in the state space as well, so that the basis for comparing a theory to the world is a mathematical relation between data points and model curves in the space.

The two approaches share what I will call the *basic semantic view*. The basic semantic view, in contrast to the logical positivists' view, holds that scientific theories are best understood as extralinguistic entities that may be described in language in various ways. Thus, a theory is a structure or set of structures which satisfies (in the logician's sense) its linguistic formulations. To give a theory directly, one must give these structures. So, as van Fraassen says, to present a theory in science is

> to specify a family of structures, its *models*; and secondly, to specify certain parts of these models… as candidates for the direct representation of observable phenomena.[2]

This basic account needs fleshing out if it is to serve as a description of what it means to put forward a scientific theory, because it doesn't say anything about *how* models, or parts of models, are supposed to directly represent observable phenomena. The question of how scientific theories are to be verified—how they can be compared to empirical observations, and accepted as representing those observations faithfully—was a stumbling block for the logical positivists, and the semantic view must give an explicit account of this process if it is to be an improvement over its predecessors.

A natural way of extending the basic semantic view to meet this challenge is to claim that scientists relate models to empirical data through *statistical fit*. This approach has the advantage of being able to account for how theories describe the world *more or less faithfully*, and doesn't face a dilemma of needing to classify all theories as "verified" or "meaningless," as the positivist view did. It also harmonizes with actual scientific practice: there is no doubt that at least some important part of scientific activity involves using statistics to confirm hypotheses about a domain of study. (By contrast, it is less clear that scientists do anything like construct the hierarchy of models in the set-theoretical way that Suppes suggests, or that such a hierarchy could capture the varying degrees of success that theories have in describing empirical data.) The claim that the model-data relation is one of statistical fit is the defining property of the state space approach's extension of the basic semantic view, and the primary reason why it is both intuitively appealing and worth scrutinizing.

In what follows, I will be concerned with the additional claims beyond the basic semantic view that are held by advocates of the state space approach. I will then discuss a particular example from network theory, the α-model of Duncan Watts, and the "small world problem" it was designed to help elucidate. The α-model, I contend, is a challenging case for the state space approach, because it seems not to be idealized in a way other than state space advocates think it should be, and because the way Watts compares it to the world is not through statistical fit of data.

2 van Fraassen, Bas C. *The Scientific Image*. Oxford: Clarendon Press, 64.

The Semantic View of Theories

Under the semantic view, a theory is to be thought of as independent of how it is specified in language. We identify a theory by directly picking out the class of models or structures that satisfy its axioms. The notion that a model 'satisfies' a linguistic formulation of a theory is borrowed from logic: it is the notion that structures *make true* or *realize* the meanings of linguistic sentences and symbols. For example, the sentence "This structure is a dense linear ordering" is satisfied by $\mathbb{R}$, the structure of the real numbers, but not by $\mathbb{N}$, the structure of the natural numbers. If T is a theory of dense linear orderings, then $\mathbb{R}$ is a model of T and $\mathbb{N}$ is not.

Proponents of the semantic view extend this notion of "satisfaction" or "being a model of" beyond purely formal and mathematical languages to the language of science. For example, both a universe in which the center of gravity has a velocity of zero and a universe in which the center of gravity has a velocity $v > 0$ (all other things being equal) are models of classical mechanics, since Newton's axioms are satisfied in both. The domain or "intended scope" of the theory of classical mechanics is *everything that has mass*; so any physical system in which the laws of classical mechanics hold is a model satisfying the theory, whether it is as large as the galaxy or as small as an apple falling to the ground. Other scientific theories have different domains, and therefore different models. In biology, for example, populations of a predator and its prey whose densities fluctuate according to Volterra's equations are a structure satisifying Volterra's theory.

It's important to notice that we are still employing the term "model" here in the logician's sense, wherein a *structure* is a model of a *theory*. Models are concrete and particular, while theories, as linguistic entities, are abstract and may be interpreted differently in different domains. It's not important that the objects or relations in a model actually exist; it's only important that, if they do exist, then they satisfy the theory asserted about them. There may not be any real predator-prey populations or physical systems satisfying Volterra's or Newton's equations exactly, but this does not prevent us from describing a variety of situations in which either theory *would* be true. This is a very different use of the word "model" than is typically found in science. The scientific meaning of "model" is almost the opposite of its meaning in logic: biologists would say that Volterra's equations are a mathematical model of real fluctuations in the population densities of predators and prey under certain conditions. In order not to confuse the two, some semantic view theorists have introduced different terminology for the scientific notion of "model." Van Fraassen, for example, calls the scientific notion a "model-type." He explains:

> In the scientists' use, 'model' denotes what I would call a model-type. Whenever certain parameters are left unspecified in the description of a structure, it would be more accurate to say… that we described a structure-type. Nevertheless, the usages of 'model' in meta-mathematics and in the sciences are not as far apart as has sometimes been said. I will continue to use the word 'model' to refer to specific structures, in which all relevant parameters have specific values.[3]

The idea here is that scientists often think of models as mathematical structures that capture what is common to many particular systems, while logicians think of those particular systems as models. So, while a physicist will say that

$$x = \sin(\omega t) + B\cos(\omega t) \quad (1)$$

models the position x of a harmonic oscillator over time, a logician will think

3 Ibid, 44.

that the values of A, B, and ω must be specified before the equation actually represents a model of the theory of harmonic oscillators. Part of the reason for making this distinction is that proponents of the semantic view like van Fraassen believe that what makes a model *scientific*, and not simply mathematical, is that it (or some part of it) is capable of being empirically confirmed. Since (1) can only be confirmed after A, B, and ω have been given, it's this specified version that is the true model of the theory.

So far, I have been characterizing what I earlier called the "basic semantic view," the view that theories are to be understood as collections of models. Following van Fraassen, we can summarize the basic semantic view as making two claims about how a scientific theory T should be understood:

Semantic claim: T presents a family of structures, its models, which satisfy the relations asserted by T in a certain domain. Understanding T involves understanding these models, independently of how they are described in language.

Empirical claim: T specifies properties of its models as candidates for the direct representation of observable phenomena.

One of the main goals of a semantic approach to the analysis of scientific theories is to avoid the pitfalls that accompanied the syntactic approach of the logical positivists. On the standard positivist account, theoretical claims are syntactically reducible to claims in a "pure observation language" through the application of the theory's internal logical calculus and the theory's bridge laws, which define basic theoretical concepts in terms of purely observational concepts.[4] Under this approach, the way that theoretical claims (reduced to claims using only observation concepts) are compared to the world must essentially be taken as primitive: after a sufficient reduction takes place, the observation concepts in a claim are supposed to be simple enough that we can simply *tell* whether or not they obtain by observing the world. The whole positivist project rests on the idea that scientific theories, claims, and concepts are meaningful because they are *verifiable*, but no one has been able to satisfactorily explain what 'verification' is, and how to resolve disputes about whether or not observations verify scientific claims.

There is thus a gap in the positivists' account of the relation between theories and the world, which can only be closed by specifying how scientific theories are evaluated in light of empirical data. Any account which improves on the positivist approach must not take verification' as a primitive; it must say how the world and our observations of it can satisfy a scientific theory, or discredit it. The basic semantic view does not achieve this on its own: the claim that a theory specifies properties of its models as "candidates for the direct representation of observable phenomena" does not go far enough in describing how that representation is achieved, how scientists determine whether an observation confirms a theory, how they deal with observational error, or how they evaluate competing theories in the same domain. The basic semantic view must therefore be extended if it is to be a credible account of how scientific theories are structured.

As I have already mentioned, the state space's approach to extending the basic

4 The canonical example is that of claims in psychology about pain, which is a 'theoretical concept' because it allegedly cannot be observed directly. Rather, we infer that a psychological subject is in pain because, for example, we observe it screaming, writhing in a way that suggests agony, trying to escape the stimulus, etc. Hence, we can in principle reduce claims about pain to claims about the observation concepts 'screaming,' 'writhing,' 'escaping,' and so forth via a bridge law that states "x is in pain if and only if x is screaming, or x is writhing, or x is trying to escape...

semantic view is to argue that the relation between scientific theories and the world is just the relation of statistical fit between the theory's models and empirical data. This approach is appealing because it provides a way to talk about the nuances of evaluating theories in light of data: data can fit a model, or a class of models, more or less well, and with some precisely-determinable error. Describing scientific judgments about a theory's value is a matter of appealing to these concepts, which scientists often use themselves. But analyzing scientific theories and their relation to empirical data in terms of statistical fit entails the claim that, for a given theory, it is possible to construct a mathematical space where a theory's models (or parts of them) and observations can be plotted, because statistical methods can only be applied in such a space. Proponents of the state space approach need to provide some philosophical infrastructure to justify this claim. This infrastructure constitutes their extension of the basic semantic view.

How does the state space approach get to a notion of statistical fit from the basic semantic view? Frederick Suppe provides an exposition in *The Structure of Scientific Theories*.[5] First, he says, it's important to recognize that any scientific theory has a domain of phenomena known as its *intended scope*: that part of the world which the theory intends to describe and explain. It would be extremely difficult, and not very useful, for a theory to try to describe each of the phenomena in its intended scope separately; rather, a scientific theory "abstracts certain parameters from the phenomena and attempts to describe [all] the phenomena in terms of just these abstracted parameters."[6] This is what is typically meant when philosophers of science say that "theories organize phenomena": they mean that theories describe many phenomena in some domain in terms of just a few concepts or parameters, which gives us a mechanism for saying how distinct phenomena are related (the same concepts apply to both) and how they are different (different concepts apply to them, or different values of the same parameters). So far, then, Suppe is presenting a fairly standard view.

What he says next, however, is slightly more controversial:

> In effect the theory assumes that only the selected parameters exert an influence on the phenomena and thus that these parameters are uninfluenced by any other parameters in the phenomena. As such the theory assumes that the phenomena are *isolated* systems under the influence of just the selected parameters.[7]

I suspect that this assertion might draw criticism from scientists, particularly those working in fields where the phenomena are explicitly recognized as being influenced by more parameters than the theory uses or can discover. This is true in ecology, for example, where the natural systems under study are *necessarily* unisolated from the environments in which they occur. Still, it seems reasonable to assume that, for pragmatic reasons, scientists often make some kind of simplifying assumptions when constructing a theory. These might come in the form of an assumed-isolated system, or they might be something weaker, like an "other things being equal" clause, or the agreement to lump unknown parameters into an error-term in their equations. At any rate, it's probably true that no theory is perfect, and that one reason many theories are imperfect is that enumerating and understanding all the parameters influencing their intended domains is a long and difficult task.

5 Suppe, Frederick. *The Structure of Scientific Theories*. Urbana: University of Illinois Press.
6 Ibid, 223.
7 Ibid, 223.

The conclusion that the state space view draws from the fact that theories use an incomplete set of parameters to describe the phenomena in their intended domains is that the systems described by theories are *idealizations*. Suppe and others (somewhat confusingly) refer to these idealized models as "physical systems." I will call them "physical models" in order to avoid confusing them with *actual* physical systems, and to highlight the fact that they are just the structures that the semantic view thinks are so important to understand. An ideal harmonic oscillator, whose position is exactly described by equation (1) once the parameters have been specified, is an example of the physical models that Suppe has in mind. The physical models of a theory are described completely in terms of the theory's selected parameters for describing phenomena in its domain. Because these parameters are never enough to completely characterize actual, unisolated phenomena in the theory's domain, the physical models are idealizations of those phenomena; they are, on Suppe's view, what the phenomena *would have been* if those phenomena were free of the influence of outside parameters.[8] [9]

Once scientists have enumerated the parameters of a theory and constructed the physical model-types they believe to characterize what the phenomena would be in isolation from all other parameters, the state space approach is ready to begin its analysis of the theory. A *state* is a specification of a value for each of the theory's parameters. The states allowed by the theory are just those that fit into its physical model-types, i.e., its mathematical laws. The theory of ideal gases, for example, has the model-type:

$$PV = nRT \quad (2)$$

Thus, states in this theory are ordered tuples of the form $\langle g_P, g_V, g_n, g_T \rangle$: an ideal gas g's pressure, volume, number of moles, and absolute temperature. (R is a constant.) Tuples not satisfying equation (2)—for example, those in which $g_P\, g_V > g_n \mathrm{R} g_T$—are not allowed states in the theory.

If a theory has n parameters, we can construct an n-dimensional *state space* consisting of all the possible (both allowable and disallowed) tuples under some ordering. That is, we let the dimensions of the state space be the parameters of the theory, and let the points in the space be the possible states arranged along each dimension according to the ordering. In the four-dimensional ideal gas space, for example, the point $\langle 2_P, 2_V, 1_n, 300_T \rangle$ is 'further out' along each dimension than the point $\langle 1_P, 1_V, 0.5_n, 150_T \rangle$. (Depending on the units used, of course, neither of these points is necessarily an allowed state in the theory of ideal gases.) Most often, the ordering of each dimension is the usual ordering of the real numbers, but it need not be; so long as the possible values for each parameter can be ordered in *some* way, they can be arranged in a state space.

Since the laws of a theory impose contstraints on which states are allowable among all the possible states, and the physical models of the theory are those structures that satisfy its laws, the physical models cover all of the allowed states in the state space; the other points are not allowed by the theory, so the theory has no models to occupy them. Quite often (though not necessarily), the physi-

8 Ibid, 224.

9 The notion of 'idealization' has not received as much attention in the literature as perhaps it should. It seems likely that systems can be idealized during the construction of models in a variety of different ways, depending on the intentions of the scientist building the model. To admit the influence of unknown parameters is one kind of idealization, for example, but the exclusion of mathematical terms containing only known parameters for the purpose of mathematical simplicity is another. Since it is my goal here to present the state space approach to scientific theories, and not to provide a taxonomy of the kinds of idealization, I have chosen to simply present Suppe's view and leave the notion otherwise unanalyzed.

cal models of a theory are dynamic, and have a value for each parameter through some length of time. We thus envision physical models as *trajectories through the state space*, that is, as curves parameterized by time drawn in n dimensions. If the models are static, every model is identified with exactly one point in the space.

We can summarize the discussion of the state space view so far with the following additional claims about a theory T:

Idealization claim: T describes phenomena in its intended domain in terms of an incomplete set of parameters, but its models are described completely in terms of those parameters; therefore, the models are idealized representations of actual phenomena.

State space claim: T can be represented as a set of allowed states in the n-dimensional state space $S(T)$ whose dimensions are the parameters of the theory. The models of T cover the allowed states in $S(T)$, and none of the disallowed states. If the models are dynamic, they can be seen as time-parameterized curves through the state space.

These two claims provide the foundation for getting to the ultimate goal of the state space approach: the further claim that models and data are related by some concept of statistical fit. The idealization claim says that a theory's models and the phenomena of its intended scope share some properties (even though they probably do not share *all* properties), so it is reasonable to compare them with respect to those properties. The state space claim says that there is a mathematical space whose dimensions correspond to those shared properties, which are the parameters of the theory. It is just such a space which is needed when one's goal is to calculate the fit of empirical data to a theory's models.

We can now proceed to discuss the state space approach's concept of "statistical fit" with some precision. The idea is that, given the parameters and models of a theory, we can compare the theory to the world by making observations of the natural systems the models are supposed to describe, plotting those data in the state space, and seeing how well the models and data overlap or coincide. *Fit* is the mathematical relation, or class of mathematical relations, that escribes the degree of success with which a model overlaps data in the state space. There are different ways to define the mathematical fit of data to a model curve that are appropriate in different situations. Thus, fit is a flexible enough notion to capture the different types and strengths of verification. We can say that data fit a model more or less well; we don't have to make a binary choice as to whether or not they verify a scientific theory.

Van Fraassen and Lloyd both use the term "isomorphism" to describe the mathematical relation of fit. Lloyd describes the model-world relation this way:

> Empirical claims are made about relations between models and natural systems; a natural system is described by a model when the model is isomorphic in certain respects to the natural system.[10]

Similarly, van Fraassen says that for a theory to be empirically adequate[11] is for it to have some model such that "all actual appearances are identifiable with (isomorphic to)" a part of that model, where "isomorphism is of course total identity

10 Lloyd, Elisabeth A. *The Stucture and Confirmation of Evolutionary Theory*. Princeton: Princeton University Press, 72.

11 Empirical adequacy is van Fraassen's anti-realist notion of what science aims for its theories to be. It is weaker than the realist notion of correspondence truth, but stronger than the standard notion of being 'well-confirmed' by past observations. Since van Fraassen claims that the semantic view is neutral on the question of realism, I take it that the notion of isomorphism he gives here is meant to apply to more than just empirical adequacy; it is meant to apply to all sorts of formulations of why we believe a theory is a good one. See van Fraassen for more.

of structure."[12]

In a strict sense, an isomorphism is a bijective correspondence between two sets that preserves relations on those sets. We might therefore interpret the state space view as holding that actual data points must be in bijective correspondence with a set of points belonging to one model of a theory for the theory to have adequately described the phenomena. Moreover, if there is any relation between the data points, such as succession in time, those relations must be preserved when the points are mapped onto the appropriate model: earlier data samples must map onto earlier points on the model-curve in the state space, and so forth.

This would be a very strong claim for the state space view to make—so strong, in fact, that it would probably be untenable. The relation of isomorphism, taken in this strict sense, is both too strong and too weak to be a good candidate for a relation between models and data. It's too strong because it would be too difficult to obtain, most of the time: it's extremely unlikely in most situations that a reasonable number of observations will be in perfect correspondence with a set of points on one of the model-curves in the state space. Each additional observation increases the chance that some data point's relations to the other points will not be preserved by any map onto a curve in the space. What's more, this interpretation of "isomorphism" ignores one of the main tenets of the state space view: that models are idealizations, so they won't fit any data set *perfectly*. It's for this reason that we brought in the notion of fit in the first place.

On the other hand, the strict interpretation of isomorphism is too weak because isomorphism is a symmetric relation: if A is isomorphic to B, then B is isomorphic to A. This is a problem because we think the relation of 'modeling' is *asymmetric*: if a structure A is a model of phenomenon B, it's not also supposed to be the case that "B models A," at least not in the usual sense. Moreover, two models of a theory can be isomorphic to each other without our being able to see that they are models of any real phenomena at all, just as two pictures of the same object can be isomorphic without either being a representation of the other. (Rather, they are both representations of the object itself.) Hence, strict isomorphism is too weak a notion to pick out the relation between models and data.[13]

It's possible that the state space view has the resources to deal with one or both of these objections, but these objections are really directed at a straw man. I think both van Fraassen and Lloyd mean "isomorphism" in a weaker sense than the strict one I sketched out above, despite the bold claims with which they introduce the notion. In other parts of their work, both seem content to leave the business of relating models and data to traditional statistical methods. Van Fraassen, for example, says that "the measurement of how well a probabilistic model fits the data gathered… is a subject already of extensive study in statistics"[14]; presumably, this is true of deterministic models as well. Lloyd similarly asserts that "fit can be evaluated by determining the fit of one curve (the model trajectory or coexistence conditions) to another (taken from the natural system); ordinary statistical techniques of evaluating curve-fitting are used."[15] [16] All of this seems more in line with the state space view's general slogan that "philosophy of science should use mathematics, not meta-mathematics" than the view that the model-data relation is to be borrowed from logic.

12 van Fraassen, 43-45.

13 This objection is due to Roman Frigg. For more on this topic, see Roman Frigg.

14 van Fraassen, 194.

15 Lloyd, 147.

16 Lloyd also gestures toward Suppes' set-theoretic approach as a way of relating models and data. It may be, she thinks, that a series of different "fits" are required to compare the models of a theory to data—for example, when there is no observable analog in the data for a concept used to define the theoretical models (Lloyd, 146). Since I am not dealing with Suppes' approach here, I leave the exploration of this idea to the reader, but it is important to note that the two approaches to the semantic view are not entirely divergent, and that one may make up for the other's faults.

The most judicious interpretation of the state space approach, then, is that the notion of fit between models and data in the state space is basically the same as the series of concepts and mathematical relations that a statistician uses to describe the fit of a collection of data to a curve. This is already a broad and flexible notion; there are many different types of statistical fit, and different tests for determining the degree of a particular kind of fit. Moreover, it is a notion that is easily extended: if in some circumstances statistical fit does not apply or is inadequate, different types of fit that are similar in spirit can be defined. For example, when the models of a theory are not dynamic and hence don't correspond to curved trajectories in the state space but only to points, we might think of "goodness of fit" between a collection of a collection of data-points and a model-point as being inversely related to the volume of the n-dimensional solid required to enclose both the model-point and the data-points in the state space.

The final claim of the state space approach, then, can be summarized as follows:

Model-data relation claim: The basis for comparing a theory T to the world is a mathematical relation between the models of T and the data from observations of phenomena in T's domain. This mathematical relation consists in the statistical fit of the data to T's models, thought of as curves in its state space, or in some similar relation appropriate to T.

As we will see, this claim and the others will figure into the state space view's account of the theory that I now turn to: the Watts α-model.

α-Model and Network Theory

I shall now discuss a theory that I believe demonstrates the usefulness and appeal of the basic semantic view, but isn't satisfactorily described by the state space view. This theory is known as the "α-model," and was presented by Duncan Watts in *Small Worlds: The Dynamics of Networks Between Order and Randomness*.[17] The α-model is an algorithm for generating undirected graphs with properties similar to those of real social networks. It was proposed as an initial (hence "α") approach to answering questions about what Watts calls the "small world phenomenon."[18]

The small world phenomenon is the generalized version of the colloquial observation that "everyone is connected within six degrees of separation." The idea is that every person belongs to a social network, the network of acquaintanceships. These networks overlap, forming a global web in which it is possible to choose any two people at random and find a series of acquaintances that connect them. If we choose persons A and D, for example, and A knows B, B knows C, and C knows D, then A and D are somehow "connected," even though they may not know each other, or even be aware of each other's existence. The "six degress of

17 Watts, Duncan J. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton: Princeton University Press.

18 Though Watts thinks of the α-model as a mathematical model, I will throughout the rest of this paper think of and refer to it as a "theory." The reason for this is that α-model consists of a set of rules for picking out what advocates of the semantic view would call its models: a class of undirected graphs. Though I will not use the term "α-theory" in order to remain consistent with Watts, it would be more appropriate to think of the graphs it generates, and not the α-model itself, as the models of the theory.

separation"claim is that any two people, chosen at random in the global acquaintanceship network, can be connected in this way with six or fewer intermediate acquaintances in the chain. In this sense, the world of human social relations is small, even though it has billions of members.

This claim about six degrees of separation, if it is true, is a surprising result to most people. The reason is that acquaintanceships are highly *localized* in that most of any one person's acquaintances are confined to a small geographic area, and in that many of those acquaintances are also acquaintances of each other. Given that every person can only know a relatively tiny fraction of all the people in the world, and that many of the links from one's acquaintances won't reach outside one's own social circle, the idea that there are fewer than six links required to reach *anyone* else from a given starting point seems unlikely.

There are also many other types of networks in the world, of course: computer networks, neural networks, the network of genetic lineage. The question arises as to whether any of these might also be "small worlds," in the sense just described: whether, though they have large numbers of members, any member can be reached from any other in a small number of steps along the network's connections. The questions of whether small worlds exist, under what conditions they arise, and how to describe their properties, are the focus of Watts' book. He formulates the problem this way:

> Assuming that a network can be represented by nothing more than the connections existing between its members and treating all such connections as equal and symmetric, a broad class of networks can be defined, ranging from highly ordered to highly random. The question then is *Does the Small-World Phenomenon arise at some point in the transition from order to disorder, and if so, what is responsible for it?*[19]

Watts is particularly interested in the "transition from order to disorder" because it seems, *prima facie*, that the topologies of many real, natural networks like the acquaintanceship network lie somewhere between those of highly-structured artificial networks and those of random networks, because connections are neither centrally planned nor formed totally independently of those already in place. The α-model was designed to take this feature of natural networks into account, and to investigate the small world phenomenon in a rigorous and mathematical way. Before I can describe the details of the model, however, a short exposition of concepts from graph theory is in order.

A network theory primer

The most fundamental concept required for understanding the α-model is the concept of an *undirected graph*. An undirected graph is a mathematical structure that consists of a set of points ("vertices") and a two-place symmetric *edge relation* describing the connections ("edges") between those points. Formally, we denote a graph by $G(E,V)$, where V is the set $\{v_i\}$ of vertices in g, and the edge relation E is the set of pairs $\{\langle v_i, v_j\rangle : \{v_i$ is connected to $v_j\}$. This is the sort of structure that Watts has in mind when he says that a network can be represented by "nothing more than the connections existing between its members." To represent a network with an undirected graph, you need only know which vertices are connected to which others; you don't need to know anything about what those connections are like, what sort of objects the vertices are, or any other information that might be relevant to describing a network in other contexts.

19 Ibid, 24.

Given an undirected graph g, a *path* between two points v_1 and v_2 in g is a set of edges that connect vertices between v_1 and v_2; it's a set of edges you could "walk along" to arrive at v_2 from v_1 if you could only step along the edges of the graph. For any two vertices in g, there is a *shortest path* of L edges between them. For example, if v_1 is connected to v_2, v_2 is connected to v_3, v_3 is connected to v_4, and v_4 isconnected to v_5, there is a path of length $l = 4$ between v_1 and v_5, so the shortest path between v_1 and v_5 is not longer than 4 edges. The path from a point to itself has length 0. If no path exists between v_1 and v_2, we say the path between them has *infinite length*. If there are any paths of infinite length in g, it is *disconnected*; if all paths are finite, then g is *connected*, meaning that any point can be reached from any other by walking along the edges of the graph. (Note that we are assuming here that the set of vertices V is finite. This is a reasonable assumption for graphs intended to represent real networks, but it is by no means mathematically necessary. In infinite graphs, however, the correlation between finite path lengths and connectivity breaks down.)

Given these definitions, we can define two macro-level properties of finite undirected graphs that are of special interest in the α-model. The first of these is the *characteristic path length*: the characteristic path length $L(G)$ of a graph is the median of the means of the shortest path lengths from each vertex v_i to all the others. That is, for each vertex v_i we calculate the length of the shortest path $d(v_i, v_j)$ between v_i and every other vertex v_j. We calculate the mean length of these paths D_{v_i}, which in a graph of n vertices is:

$$D_{v_i} = \frac{\sum_{i \neq j} d(v_i, v_j)}{n-1}$$

Then we calculate the median of these D_{v_i} to obtain the characteristic path length $L(G)$:

$$L(G) = median(\{D_{v_i} : v_i \in V\}) \quad (3)$$

Intuitively, the characteristic path length is the average number of steps required to walk along the edges of the graph from a typical vertex to another vertex chosen at random; it measures how 'close' the vertices of the graph are in the absence of an underlying spatial metric.

The second macro property of graphs that we are interested in with respect to the α-model is the *clustering coefficient* $\gamma(G)$ of a graph. To define this property, we first define the *neighborhood* of a vertex V to be the set of points $\Gamma(v)$ in g that are connected to V: $\Gamma(v) = \{ v_j \in V : \langle v, v_j\rangle \in E\}$. Note that by convention, $v \notin \Gamma(v)$. The *degree* k_v of a vertex V is the number of neighbors it has: $k_v = |\Gamma(v)|$. The *clustering around a vertex* γ_v is the total number of edges between v's neighbors divided by the total number of edges that are possible amongst those neighbors:

$$\gamma_v = \frac{\text{total number of edges in } \Gamma(v)}{\text{total possible edges in } \Gamma(v)} = \frac{|\{\langle v_i, v_j\rangle : v_i, v_j \in \Gamma(v)\}|}{\binom{k_v}{2}}$$

The idea is that the clustering around a vertex v captures the notion of how closely-knit v's 'social circle' is, if we imagine that the edge relation in G represents friendship between two vertices. If $\gamma_v = 1$, then all of v's neighbors are friends with each other in addition to being friends with v. If $\gamma_v = 0$, on the other hand, then none of v's neighbors are friends with each other; in the absence of v, the paths between them have lengths of at least 2.

Given this notion of clustering around a vertex, we can finally define the clustering coefficient $\gamma(G)$ of the graph as the average over the clustering around all vertices. If G has n vertices, then

$$\gamma(G) = \frac{\sum_{v \in V} \gamma_v}{n}$$

The clustering coefficient of the graph captures for the whole graph what the clustering around a vertex captures on a local level: how 'tightly-knit' or locally-clustered the graph is on average. When $\gamma(G) = 1$, G consists entirely of one or more *completely connected components*: a series of subgraphs in which every vertex is connected to every other, but no vertex is connected to any others in other subgraphs. When $\gamma(G) = 0$, by contrast, no neighbor of a vertex v is connected to any other vertex in $\gamma(v)$. Such a graph is usually very *sparse*; that is, the total number of edges in the graph is much smaller than the total number of edges that are possible: $|E| \ll \binom{n}{2} \gg$. It is possible, however, for a relatively sparse graph to still have a high clustering coefficient: if G consists of many completely connected but disjoint subgraphs, for example, then $\gamma(G)$ will be equal to 1, but G will be sparse because each vertex is connected to only a small fraction of the total number of vertices it *could* be connected to.

When a graph G has a high clustering coefficient and a low characteristic path length simultaneously, and it is also relatively sparse, G is intuitively a 'small world': though connections in the graph are highly localized because $\gamma(G)$ is high, the 'degree of separation' between any two vertices is low because $L(G)$ is low. G should also be sparse, because it's trivial that very dense graphs can have high $\gamma(G)$ and low $L(G)$: it's no surprise that the world is small if everyone knows almost everyone else. It is therefore the questions of when and how these properties co-occur that the Watts α-model attempts to answer.

The α-model and its parameters

How can networks that are highly clustered, as well as relatively sparse, exhibit short characteristic path lengths? For Watts, the question is motivated by what we know about random graphs and about real social networks. Random graphs (that is, undirected graphs of the type described above, generated by forming edges on a set of n vertices by choosing from all the possible edges $\langle v_i, v_j \rangle$ uniformly at random) have a low characteristic path length, but seem to be poor models of real social structure, since, in particular, it's no more likely for two vertices in a random graph to be connected to each other if they have a common neighbor than if they do not. In reality, by contrast, many of my friends know each other, as do many of theirs, and so on: the clustering coefficient of the acquaintanceship network should be much higher than the typical clustering coefficient of a random graph. On the other hand, some empirical work[20] suggests that the characteristic path length of the acquaintanceship graph is relatively short. Thus, it seems that real social networks like the acquaintance graph have properties in common with both random graphs and highly-clustered graphs with lots of local structure, but are identical with neither.

The Watts α-model is an algorithm for generating graphs that can have both a high clustering coefficient and a low characteristic path length. It builds a graph between a chosen number of vertices n in a stepwise fashion, adding a single edge to the graph at each iteration of the algorithm, and terminating after there are

20 Watts credits Stanley Milgram with beginning an empirical investigation of the small-world phenomenon in the 1960s (Watts, 18). See, for example, Stanley Milgram.

enough edges in the graph that its average degree k reaches some pre-defined value. The resulting graphs have values of γ and L that are determined by the parameters of the algorithm.

The algorithm proceeds by first visiting each vertex i in the graph in turn and calculating its "propensity" $R_{\{i,j\}}$ to connect to every other vertex j in the following way:

$$R_{i,j} = \begin{cases} 1 & \text{if } m_{i,j} \geq k \\ \left[\frac{m_{i,j}}{k}\right]^{\alpha} & \text{if } k > m_{i,j} > 0 \\ p & \text{if } m_{i,j} = 0 \end{cases}$$

The parameters k, p and α and the variable $m_{i,j}$ in this equation have the following meanings:

- k is the average degree of the graph (that is, the mean of k_v over all vertices v), which, like n, is specified in advance
- $m_{i,j}$ is the number of neighbors that i and j already share on the current iteration of the algorithm
- p is a baseline random probability that i and j will connect, even if they have no neighbors in common; $p \ll \binom{n}{2}^{-1}$
- α is a "tunable parameter" defined on $[0, \infty)$

Intuitively, $R_{\{i,j\}}$ formalizes the notion that connections form in social networks between two vertices i and j with different probabilities, depending on how many acquaintances i and j already share. i and j have some random, baseline probability p of meeting, even if they don't have any common friends. At the other extreme, if i and j share more acquaintances than most people know, they have too many common friends *not* to be connected to each other. How connections form between vertices between these two cases is controlled by α.

Once all the $R_{\{i,j\}}$ have been calculated, they are normalized to the unit interval: each vertex j is assigned some half-open interval on $[0, 1)$, the width of which is proportional to the fraction with which i has a propensity to connect to j among all other vertices:

$$\text{width of } j\text{'s interval} \propto \frac{R_{i,j}}{\sum_{i \neq v} R_{i,v}}$$

Note that j's interval is disjoint to the intervals assigned to every $v \neq j$. A random number is then generated on $[0, 1)$, which must fall into one of these intervals—say, the interval for vertex V. Then an edge is created between i and V. This process repeats until the chosen value of k is realized by the graph.[21]

By fixing n and k in advance, we can guarantee that the resulting graph will be sparse; thus, we are interested in the effect of the other parameters on the properties of clustering and characteristic path length.

The probability p that a vertex i will connect to a vertex j with which it shares no neighbors is what allows the algorithm to get up and running. Starting from a set of vertices with no edges between them, p guarantees that every $R_{\{i,j\}}$ is non-zero, so that the intervals of $[0, 1)$ corresponding to the vertices j actually have a width. Without this condition, no edges would ever be formed. However, p contributes to the 'random character' of the resulting graph: the larger p is, the better chance there is of i connecting to a vertex j_1 with which it shares no neighbors as compared to the chance that i will connect to a vertex j_2 with which it shares one or more neighbors. For this reason, p is kept very small, so that the amount of randomness

21 Watts, 46-47.

in the graph generated is almost completely determined by the central parameter, α.

The parameter α of this algorithm is used to vary the amount of randomness in the resulting graph. When $\alpha = 0$ (a circumstance Watts refers to as the "Caveman world"), new connections are formed at each iteration almost entirely on the basis of the existing edges, since $R_{\{i,\,j\}} = 1$ for all vertices i and j that share at least one neighbor, and $R_{\{i,\,j\}}$ is small (equal to p) for vertices that have no common neighbors. The resulting graph is highly clustered, consisting mostly of isolated 'caves' of vertices that are all connected to each other, but not to other vertices outside the group. On the other hand, as α approaches infinity (a circumstance Watts calls the "Solaria world"), new edges form almost entirely at random, since the $R_{\{i,\,j\}}$ approach 0 for vertices that share at least one but less than k neighbors. Hence, the other cases (where $m_{i,j} \geq k$ or $m_{i,j} = 0$) become relatively more important in determining the structure of the graph. Since it is extremely rare for two vertices to share more than k neighbors, most edges form with probability p, and the resulting graph turns out to be mostly random in its structure.

Another important parameter in determining the outcome of this construction algorithm is the choice of what Watts calls a "substrate," an existing set of edges between the vertices that serve as an input to the algorithm. The substrate can have a strong effect on the structure of the output graph. If the substrate is empty, for example, then running the algorithm when $\alpha = 0$ often results in the disconnected graph of isolated 'caves' mentioned above. The characteristic path length L of such a graph is infinite by definition. To avoid this problem, Watts chose to use a ring substrate (i.e., a graph in which every vertex has exactly two neighbors), so that the resulting graph would always be connected and therefore have a finite value of L, even at low values of α. This aids in the comparison of graphs across a range of values for α, but it's a significant fix-up that reduces the size of the class of graphs that the algorithm could construct, so Watts spends a fair amount of time justifying his choice. The use of other substrates have varying effects depending on their topologies; Watts chose a ring substrate mostly because it had the *smallest* effect on the properties of clustering and characteristic path length.[22]

Conclusions from the α-model

Does the α-model produce graphs that have the desired properties of both high clustering and low characteristic path length? To make a long story short: it does. By running the algorithm many times at specified values of n, k and p and averaging the properties of $\gamma(G)$ and $L(G)$ over the resulting graphs, Watts found that as α increases, both the path length and clustering of the graphs increase briefly, then drop off sharply. The path length, however, consistently peaks and then falls *before* the clustering coefficient, so that there is a class of graphs in which $\gamma(G)$ remains high while $L(G)$ is very low. (For graphs of 1,000 vertices with $k = 10$, for example, this happens when α is between 5 and 10.)[23]

The α-model therefore demonstrates that, in the abstract, small world graphs do exist. This is the most important conclusion that can be derived from analysis of the outputs of the model: the "Small World Phenomenon," as Watts calls it, does arise for a certain class of graphs among all the graphs that could be generated by the α-model's construction algorithm, and it arises "in the transition from order to disorder" that occurs with increasing α. As a first step in the investigation

22 Ibid, 58-66.
23 Ibid, 52-58.

of the problems of defining, describing, and explaining small world networks, the α-model is a great success.

There are also some secondary conclusions to be derived from the α-model. Watts discusses some empirical results in connection with the 'relational graph model,' a construction algorithm based on the α-model and its successor, the β-model. He compares graphs generated by the relational graph model to three real networks for which complete data about the connections is available: the "Kevin Bacon graph," in which the vertices are the set of Hollywood actors, and the edges consist in the relation of having acted in some movie together; the "Western states power grid graph," in which the vertices are power stations and the edges are major power transmission lines in the western United States; and the "*C. Elegans* graph," in which the vertices are neurons and the edges are synaptic connections in the nervous system of the famous roundworm *Caenorhabditis elegans*. Without exhaustively reviewing the results here, I can give Watts' conclusion from them: each of these networks could be considered a small world, since they are relatively sparse, and have relatively high values of γ and low values of L; random graphs generated on the same number of vertices as each of these networks also have low L, but fail to have high γ; and graphs generated by the relational graph model tend to fare better in modeling both γ and L in these networks than the random graph model, though both fare pretty poorly in modeling clustering in the *C. Elegans* graph. It therefore seems that the relational graph model (and, by extension, the α-model) fits the available empirical data better than other mathematical models available.[24]

It is important to note, though, that Watts never directly compares graphs generated by the α-model to real networks, at least not in his presentation of the data in *Small Worlds*. Of course, this does not make the α-model any less of a scientific theory: it was proposed as a preliminary means of investigating an empirical problem, it is based on plausible assumptions about real social networks, it is mathematically rigorous, and it distinguishes as important two properties of networks than can be measured in both the models of the theory and in empirical data. By most reasonable standards, the α-model surpasses mere mathematical formalism or pseudo-science. Even an empiricist proponent of the semantic view will recognize that the α-model does indeed present a family of models, and the parameters of those models are intended as representations of observable phenomena, so it meets her criteria for a scientific theory. It's just that, as a scientific theory, the α-model was never intended to represent phenomena in the way philosophers of science often imagine theories do. I shall now turn, therefore, to seeing whether the state space view of scientific theories can adequately describe the kind of theorizing the α-model presents.

A Challenge to the State Space View

To show that the α-model presents a challenge to the state space view, I will defend the following claims in turn:

1. The α-model presents a family of structures.
2. The α-model specifies certain properties of these models as candidates for "the direct representation of observable phenomena."
3. The α-model does involve various idealizations and simplifying assumptions, but not necessarily of the kind that the state space view

24 Ibid, 139-161

envisions.

4. The relations between the models of the α-model and the world are not easily described by the notion of 'fit' in a state space.

If the first two of these claims are true, it means that the α-model meets the requirements of the basic semantic view for being a scientific theory, and so is a candidate for further description under the state space approach. If the second two are true, however, it means that the state space approach will not tell an adequate story about how the α-model functions as a theory; in this sense, the α-model presents a challenge to the state space view.

The α-model and the basic semantic view

It should be clear that the α-model does indeed present a family of structures. The structures it presents are a class of semi-random undirected graphs. Watts' use of the term "α-model" to refer to this whole class of graphs (or, more precisely, the rules by which those graphs are generated) is consistent with Van Fraassen's claim that, in scientific use, "model" typically means "model-type." The particular graphs generated by the α-model, for a specified number of vertices n, and specified values of k, p and α, are the actual (instantiated) models of the theory, on the semantic view. It's also clear that this class has quite a lot of members—at least enough so that Watts could do the statistics required to draw his conclusions—but not *every* undirected graph belongs to it.[25] Thus, the family of structures presented by the α-model is a non-trivial one: it has multiple members, so it's properly called a "family," but it's not so inclusive as to be uninteresting; an undirected graph must really have certain properties in common with other members of the class in order to be among the models presented by the α-model.

In this respect, the α-model is an almost ideal example of the semantic view's claim that theories are best understood as collections of models. Watts himself notes that the construction algorithm he gives is only one possible way of picking out the class of graphs that he intends[26]; it's really those graphs, and not the way they are constructed, that are the heart of the theory. Moreover, the models are very simple, in the sense that they are completely characterized by simple mathematical structures (the set of vertices and the edge relation on them), and they are even of the sort that logicians are familiar with. I therefore do not think any semantic view theorist should dispute claim (1).

The second claim is equally uncontroversial. The α-model specifies properties exhibited by its models as candidates for the direct representation of observable phenomena: namely, the properties of path length and of clustering. Indeed, Watts' stated purpose in constructing the model was to find a simple algorithm that would pick out a class of graphs in which these properties coexisted. The concepts by which these properties are defined are also 'observable' in real networks, on any reasonable definition of that term: given enough data about a network, it is easy to compute the number of vertices it has, its average degree, the length of a path between two vertices, and so forth. Even the positivists would be satisfied that the α-model does not employ any irreducibly theoretical concepts.

I take it that the basic semantic view's empirical claim, as I sketched it out above, is mostly meant to exclude purely mathematical theories from the domain of theories it is attempting to characterize. True scientific theories must be prepared to say something about the observable world. The empirical claim says only that the models of a scientific theory should have properties that *could* represent phenomena; it says nothing about how *well* the models must do so, and it says nothing about the intentions of the scientist who constructs or uses the theory in his investigation of an empirical problem. The α-model therefore satisfies both the semantic claim and the empirical claim, even though Watts may not have intended that it describe any real networks very faithfully, so it's a scientific theory that deserves further examination under the state space approach.

25 It would be impossible, for example, that an iteration of the α-model algorithm would produce an undirected cycle when the parameter k is greater than 2, because $k = 2$ in all undirected cycles, and the so the algorithm would not terminate until a more connected graph had been generated.}

26 Ibid, 46.

The α-model and the state space approach

How well does the α-model fit the picture of scientific theories that the state space approach favors? To answer this question, we have to see whether it satisfies the idealization claim, the state space claim, and the model-data relation claim.

I claimed in (3) above that the idealizations of the α-model are not of the sort that the state space view envisions. Recall that the models of a theory T are described completely by some set of parameters that scientists abstract from the phenomena in the intended scope of the theory. The models of T are said to be idealizations because the unisolated systems that scientists are attempting to describe and explain with T are under the influence of other, unknown factors.

The parameters of the α-model are: the number of vertices n and average degree k that must hold in the output graph; the baseline probability p that two vertices will connect on an iteration of the construction algorithm when they share no neighbors; α itself; and the substrate graph. It should be clear that n and k are the sort of parameters that are abstracted from the phenomena in the way that the state space view envisions: every network must have values for these parameters, no matter how it was generated. Probabilities are also standard parameters in scientific models, though they may or may not be abstracted in any obvious way from existing data, so I shall assume that the state space view has no problem accounting for p.

The other parameters, however, are more of a problem. α is explicitly given as a kind of mathematical knob by which the output of the construction algorithm is controlled; it is a formal construction without an intended correlate in real networks. Likewise, the substrate graph that is fed into the construction algorithm has an important effect on the structure of the graph that comes out; but the inclusion of a substrate is also more a formal property of the theory (one that had to be justified) than a parameter extracted from real data. In general, it is not possible to examine a network and decide what sort of substrate it must have been built on, for it may not have been built on any substrate at all. Already, then, the state space view's account of idealization seems not to capture what happens in the α-model.

What about the claim that the models that come out of the construction algorithm must be idealizations because the parameter list is incomplete? It's true that, with respect to the properties of clustering and characteristic path length, other parameters might influence the values of these parameters in real networks in a way not captured by the α-model. n, k, and p are certainly not enough to determine the values of γ and L for a given graph, and we know that α and the substrate are not parameters that have any real correlates, so something else not mentioned in the α-model must determine γ and L in the domain of small world networks. This certainly looks like the sort of idealization the state space view has in mind, then.

I would point out, though, that the state space view of idealization carries with it the idea that the parameters of a theory *would* be sufficient to characterize the phenomena in the theory's domain if those phenomena were free of the influence of unknown parameters. For Watts' theory, this is probably not true: there is no guarantee that even all the possible combinations of n, k, p, and α would generate the variety of γ and L found in real networks. (I exclude substrate as a parameter here because, if one is allowed to pick the substrate from all the possible undirected graphs, a substrate could be selected that already had all the desired properties, making the representation of all γ and L a trivial matter.) Moreover, I doubt that Watts would claim that his theory was idealized in this way: if his goal had been a completely faithful representation of what the properties of real networks would be in the absence of other parameters, he would not have made an unrealistic parameter like α the central parameter of his theory.

I do not deny, therefore, that the state space view has something interesting to say about idealization in the α-model, but I think the true story is more complex than even the somewhat strong notion of idealization presented by Suppe suggests. The α-model is idealized not just in the sense that its parameter list is incomplete, but also in the sense that some of its parameters do not correspond in any straightforward way to properties of phenomena in its intended domain. I maintain, then, that claim (3) above is a reasonable one.

And what of the fourth claim, that the notion of 'fit' in a state space between the structures of the α-model and observation data does not capture the ways in which the theory is compared to the world? Suppose we constructed a state space for the α-model. What would this space look like? The most natural way to build it would be to include dimensions for all the numerical parameters of the theory (that is, n, k, p, and α) and dimensions for the two properties we are interested in, L and γ.

Since the α-model was not presented as a theory intended to model the *topologies* of real networks (that is, the exact set of of edges on a specified number of vertices), we do not need to include any dimensions in the space that would encode graph topology, even though it would be *possible* to compare the topologies of real networks to the topologies of α-model graphs. Watts was really only interested in seeing how the macro-properties of clustering coefficient and path length were affected by varying α; so long as our space includes those three dimensions, then, it is complete.

We can now say that each graph G generated by the construction algorithm of the α-model corresponds to exactly one point in this space, which we can name using the vector $\langle n_G, k_G, p_G, \alpha_G, \gamma_G, L_G \rangle$. Note that this graph need not be unique: another, different graph H generated by the construction algorithm could have the same values along each dimension. Since the models of the α-model theory are not dynamic, we don't extend them to include a series of successive states in the state space—they are not trajectories but points, so the traditional notion of statistical fit of data to curves in this space won't work as a comparison of the α-model and the world. Moreover, because a point $\langle n_i, k_i, \alpha_i, \gamma_i, L_i \rangle$ does not uniquely determine an α-model graph, it would be strange to think that the fit of some collection of data to a *point* in the state space is a comparison of one model to the world. The notion of fit between real networks and the models of the α-model theory must lie elsewhere.

As I said earlier, this is not necessarily a problem for the state space view. There are many ways that we can define fit, and many ways we can extend it. I do want to emphasize, though, that an important part of the state space view's concept of fit is that data be compared to a single, specific model of the theory. Data can fit different models of the same theory more or less well, and goodness of fit might be the very relation that helps us pick out which model most faithfully represents the particular phenomena under study; but we fit the data to one model at a time. It doesn't make sense in general to say that data fit two models of the theory *simultaneously*, or that data fit the disjoint union of two models: we can draw lines describing how far a given data point is from one model curve or the other in the state space, but we can't generally draw a single line that describes "how far away it is from both."

I admit that this is a rather subtle conclusion to draw from the state space view, and that van Fraassen, Suppe, or Lloyd might easily deny the claim if presented with it directly. I think it is implicit in their discussion of how models are related to data, though, and there's no reason it shouldn't be: traditional statistical techniques are supposed to describe that relation, and those techniques are generally based on calculating the fit of a data set to one model at a time. It's just this feature of the notion of fit, though, that I think causes problems with the α-model: the most important ways in which the α-model relates to the world are not ways in which data about a single real network is compared to an α-model graph.

The best way to see this is to look at how Watts uses the α-model in his investigation of whether or not small worlds can exist in real social networks. His process, as I reconstruct it, went something like this: first, he chose values for n, k and p. Then he used the α-model's construction algorithm to generate a (reasonably large, for statistical purposes) number of graphs using a ring substrate and a particular value of α. For each of the resulting graphs, he calculated the clustering coefficient and characteristic path length, and he averaged these values over the whole collection to obtain a single point $\langle \alpha_i, \gamma_i, L_i \rangle$. He repeated this process at successive values of α until the average values of γ and L stabilized.

By plotting the points $\langle \alpha_i, \gamma_i, L_i \rangle$, Watts was able to conclude that some sparse graphs do have a high clustering coefficient and a low characteristic path length. This led him to create new algorithms (the β-model and the relational graph model) that explored the small world phenomenon in different ways. He eventually obtained data about some real networks, and attempted to fit that data to the relational graph model, which was in part derived from the α-model. He discovered that the relational graph model generated graphs whose properties tended to fit the data from the real networks better than the properties of random graphs.

If this reconstruction, which follows the arc of Watts' book, is anything like what really happened, it seems Watts was never very much concerned with how α-model graphs fit any real data. The relation between the models of the α-model and the world is mostly one of *showing small worlds to be possible*: that is, the collection of graphs generated by the α-model showed that social networks *could* simultaneously exhibit high γ and low L, given a few plausible assumptions.

If we wanted to represent this relation using a state space, we would probably select an empirical subspace of the total state space where the values of n, k, and p are fixed, as Watts did. We would show that we could generate a large number of graphs whose values for α, γ, and L were all bounded by some solid in that space, and that solid contained a region where γ was high and L was low. We would argue that the existence of this solid showed that *if* we collected data about real networks, and represented those networks as undirected graphs, those networks *could* be small worlds, because the data might fall into this region of the state space; there is more reason to think that it could than that it couldn't, given the properties of the output graphs from the α-model.

I am belaboring this point in order to make it plain that something other than 'fit' must be the relation between the α-model graphs and the world. We are not comparing real data, or potentially real data, to any particular model in the α-model state space. Rather, we are using the range of values of γ and L in the whole *collection* of models to argue that real data could fall in this range as well. We could not make this argument without reference to the whole collection; since each particular model is a graph generated by a semi-random procedure, we would have no way of using a single model to demonstrate the real possibility of small worlds, because any particular model could be an unlikely case or a mathematical fluke. If the state space view's notion of fit was the only way that scientists could compare theories to phenomena in their intended scope, the α-model would not have been able to show the important results that it did. That, I think, is reason to believe in the soundness of claim (4) above.

Conclusions

I have argued that the α-model is an empirical theory that presents a set of models whose properties are candidates for empirical confirmation, in the sense that they might be exhibited by real networks. This means that the α-model is a theory that proponents of the state space approach believe they can describe in terms of the way it idealizes its representation of phenomena and the way observation data fit its models. I have shown that the idealization inherent in the α-model is not entirely described by saying that its models are determined completely by an incomplete set of parameters: not only are the parameters of the α-model incomplete (they were never intended to be anything else); at least two of them are formal devices that have almost no interpretation in real phenomena, and could not have been abstracted from any observations. I have also shown that the α-model does have a state space, or at least we can construct one for it, and that each of its models occupies a point in this space. Nevertheless, it seems that the most important way in which the α-model relates to the world is not adequately characterized by the state space view's notion that the model-data relation is that of statistical fit. As a collection, its models demonstrated that it would not be fruitless to try and gather data about the small world phenomenon; but no data set was ever fit to one of those models, even under a broadened conception of fit.

I therefore claim that the state space approach cannot describe the α-model theory without extending its picture of how scientific theories work. Objections to this claim would probably come in one of two forms: either that I have a misinformed view about what the state space approach to theory analysis entails, or that I have been unimaginative in my description of the α-model as I applied that approach, so that my description was inadequate. I am of course open to corrections of the first sort. I do not claim to have given a complete summary of everything held by the semantic view of theories, or by the state space view in particular. I have, however, tried to give the views of multiple authors when they varied in their opinions, and to characterize the semantic view in the most judicious way possible, so I think that criticism on this point will be mild.

It's certainly true that the challenges to the state space approach that I raised in the previous section could be resolved by modifying the notions of idealization and of fit. The notion of idealization given by Suppe in his summary of the view is too limiting because it insists that the parameters of a theory be abstracted from phenomena, which makes no room for theories that have a parameter like α. This is easily fixed by enumerating the other types of idealization that operate in

scientific theories. As for fit, it would also be possible to enumerate other types (such as 'fit' between a *collection* of models and data, which would be required to describe the α-model's relation to data), until the notion was sufficiently broad to cover other types of model-data relations.

This approach would mask the real issue, however. The α-model belongs to an important class of theories that exhibit a relation between their models and the world which is not easily characterized by the notion of fit. These are theories which might be called "preliminary theories" or "theories of a problem." The α-model's purpose was to answer the questions, "Can small world networks be defined by some set of measurable properties?" and "If so, is it possible that real networks could have those properties?" By answering those questions affirmatively, the α-model laid the grounds for further empirical work. It's because the α-model was a theory of a problem in this sense, aimed at showing the *possibility* of further work rather than representing phenomena in a faithful and direct way, that it fell outside the state space view's usual notion of a scientific theory.

Parallel questions about the possibility of defining and investigating some empirical problem can be asked in many domains. For exmaple, a cognitive scientist might ask, "Can the recency and contiguity effects observed in human free memory recall be explained by a given computational algorithm, and if so, is it possible for that algorithm to be computed by a neural structure in the brain?[27] An important question concerning the development of the modern synthesis in biology was, "Can the differences in phenotypic traits of organisms be attributed to differences in their genotypic traits, and if so, is it possible to demonstrate those genotypic differences in real organisms?"[28] A physicist will wonder, "Can these phenomena be explained by positing the existence of a new kind of particle, and if so, is it possible to detect that particle?" before trying to design the accelerator which will study it. In these and similar cases, a theory was (or could be) presented which was not intended to present models which faithfully fit all the empirical data, but instead aimed at defining some empirical problem and demonstrating it to be a worthy candidate for further work. There is a good reason for such theories to be developed: scientists have limited resources, and they need some assurance that their work will be worthwhile before they plunge whole-heartedly into gathering exhaustive data and building models. These preliminary theories, or theories of a problem, can provide that assurance.

The existence of a class of scientific theories which are well-described as "theories of a problem" raises the question of whether there are other classes of scientific theories, with other types of relations to the world, which the state space view would have difficulty characterizing in terms of fit. Scientific theories bear many other relations to natural systems—and to scientists, and to other theories, and so on—and there is at least an open question about which of these other relations

27 This example comes from Howard and Kahana's *temporal context model*, which proposed an alternative mechanism for modeling these effects in free recall tasks. Their paper gave a general model that represented mental context as mathematical spaces, and recall as a well-defined algorithm which operated on those spaces. Though they do fit their model to some existing data, they note that the results are not "numerically spectacular" (Howard and Kahana, 286) and that their main purpose was to provide a simple model-type which would provide an alternative to existing theories based on random context, and serve as the basis for further work on temporal context.

28 This example, famous in the history of biology, arose from the work of Gregor Mendel (see "Versuche über Plflanzen-hybriden"), who first described the ratios of variations in phenotypic traits in pea plants in terms of the concepts of dominance and recessiveness. Mendel's results were limited in scope to what we now recognize as single-gene traits, and he could not have intended that the mathematical relations he found would faithfully fit data for other phenotypic traits in other organisms. His concepts provided the basis for further work, though, that eventually developed into the field of classical genetics.

between theories, models, data, and the world are important to have in a general characterization of scientific theories. I might, for example, compare a theory and the world by showing that a theory says certain phenomena are impossible, and look for counterexamples in empirical data. I might adopt a theory because it unifies, simplifies, or explains other existing theories, even in the absence of any new empirical evidence to support it. I might put forward a theory which contains purely qualitative descriptions of phenomena, or draw an analogy between an observation I have made and a theory from an entirely different domain. My belief or disbelief in the truth of a theory will influence the way I proceed in further work.

Proponents of the state space view have the option of claiming that the fit-relation between models and data is an *essential* property of scientific theories, and that none of these other relations or functions are essential, even though they might figure into descriptions of some or even many bodies of scientific work. In light of the α-model and similar theories of a problem, though, proponents of the state space view should ask themselves whether such essentialism is warranted.[29] If philosophers of science are seeking to understand the many facets of scientific theorizing, including how theories are developed, how they augment our knowledge of the world, and how we should think of the different relations they can bear to empirical data and to each other, then we would do well to make our descriptions of them as inclusive and as far-reaching as possible.

To my knowledge, there are no significant philosophical positions poised to replace the state space approach. If the semantic view of theories and the state space approach to it were hopelessly distorted pictures of scientific theorizing, it would signal the need for a new view to be developed. They are not. There is no doubt that the state space approach gets a lot right: scientists *do* construct models with parameters abstracted from their domains of interest, and they *do* evaluate theories in part on the basis of how their models fit empirical data. I have attempted to show here that scientists do other things as well, and that one important scientific activity involves developing theories which serve other purposes than simply faithfully representing phenomena. It is not necessary to abandon the state space view to recognize this fact; but it is necessary to be willing to supplement it in those domains of science where we find its picture incomplete.

References

van Fraassen, Bas C. *The Scientific Image*. Oxford: Clarendon Press, 1980.

Frigg, Roman. "Scientific Representation and the Semantic View of Theories". *Theoria 55* (2006), 49-65.

Howard, Marc W. and Kahana, Michael J. "A Distributed Representation of Temporal Context." *Journal of Mathematical Psychology 46* (2002), 269-299.

Lloyd, Elisabeth A. The Stucture and Confirmation of Evolutionary Theory. Princeton: Princeton University Press, 1988.

Mendel, Gregor. "Versuche über Plflanzen-hybriden" *Verhandlungen des naturforschenden Ver-eines in Brünn} IV* (1866), 3-47. [Translated by William Bateson with edits by Roger Blumberg, as part of the MendelWeb project: http://www.netspace.org./MendelWeb]

Milgram, Stanley. "The small world problem." *Psychology Today 2* (1967), 60-67.

Suppe, Frederick. *The Structure of Scientific Theories*. Urbana: University of Illinois Press, 1974.

Suppes, Patrick. "Models of Data." In *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, Edited by E. Nagel, P. Suppes, A. Tarski. New York: Springer, 1962.

Suppes, Patrick. "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences." *Synthese 12* (1960), 287-301.

Watts, Duncan J. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton: Princeton University Press, 1999.

29 Lurking in the background here is an issue that I have not considered in this paper: the question of whether or not the semantic view of theories, and the state space approach in particular, are supposed to be *descriptive* or *normative* analyses of the structure of scientific theories. It might be that some proponents of the state space view are perfectly happy to ignore "theories of a problem" because they don't think such works *should* be considered scientific theories, even if other people are content to call them that. My view is that a good philosophy of science should not be normative in this sense, however, precisely because it will cause philosophers to turn a blind eye to important and interesting types of scientific work.

AN INTERVIEW WITH AMARTYA SEN

February, 2008
Harvard University

Each year The Dualist *includes an interview with a contempoary philosopher chosen by the staff. This year, we are very pleased to have Amartya Sen answer questions posed by* The Dualist *and by the Stanford Philosophy Department.*

BIOGRAPHY NEEDED HERE
BIOGRAPHY NEEDED HERE
BIOGRAPHY NEEDED HERE

David Hills:

In Identity and Violence *you point out that each of us has many distinct identities. The most important of a person's identities aren't always these she was born with, born into, or otherwise stuck with by outside forces. The relative importance of various different identities and the terms on which these identities relate to one another in a complete human life do and perhaps must vary profoundly from person to person. Public institutions and public policies do* all *of us a serious injustice when they proceed as if single religious or cultural identity summed up* each *of us.*

This much strikes me as entirely true and of immense importance. Yet you go on to write as if each of us chooses *her identities (and the terms on which these identities will relate to one another in her own life) under* constraints *imposed by resources, personal capacities, ways others are and are not prepared to view her and treat her, etc. Certainly there are cases where the adoption and cultivation of an identity is straightforwardly a matter of free rational personal choice. Certainly one very powerful and familiar form of social scientific explanation portrays agents as making rational choices of which they are unaware, under constraints they may never explicitly conceptualize. Still, aren't some extremely important identities such that we simply* can't *acquire them (in any authentic form, at any rate) by* choosing *to acquire them? Aren't many religious and ethnic identities like this, for instance? We need to respect the capacity of human individuals to make their own choices under personally distinctive and unprecedented sets of imposed constraints. But is the first kind of respect a special case of the second? Is the consideration due us as possessors of identities a special case of the consideration due us as free rational choice-makers?*

Amartya Sen:

I see the importance of your questions. Despite the big role of "choice" in the balance of our identities, we have to see how we are constrained, in many different ways, in this choice by forces beyond our control. I have tried to clarify the fact that all choice, in every field, is invariably within some constraints (explicitly understood or implicitly recognized). The fact that a consumer cannot typically choose to buy all the contents of an entire shop (even if the person were crazy enough to want to do such a thing) because of what economists call "budget constraint" does not imply that the consumer does not have important choices to make within his or her budget constraint. So I don't see that it is problematic for my theory of identities to face your question, "aren't some extremely important identities such that we simply *can't* acquire them?" Indeed, so. The point, however, is that nevertheless there are other choices, which are both extremely important and which we can actually make, and we should not move from the basic understanding of the existence of constraints that bind all choice to the view, championed by many cultural theorists and some variants of communitarians, that there are no significant choices to be made as far as identities are concerned. That is the central issue here.

Sometimes we may not fully recognise what the limits of a specific identity might be. Let me give an example. When my young children, Indrani and Kabir, were growing up in London (we lived in London then), I told them, in answer to their questions about their ethnicity, that they were "black" (the term is easier to use in Europe than it is in America because of greater squeamishness here in using a term associated with racial disadvantage). However, since I am from India, and my late wife was an Italian, the schools in London liked classifying them, when such a classification is needed, as "brown" or "Asian" or "mixed." But I was right too, since the distinction between "black" and "white" is really what this racial classification is about, in telling you whether you are inside the racially privileged domain of being "white" or not. Everyone outside that boundary, as I saw it, was for that purpose "black." I also thought it was good for the children to acquire a broader and less divisive identity, since non-white solidarity is a good way of resisting racial inequality. Being "black" is really a social category, not pigmentational one. And "brown" is a better description of bread than of human beings, while "people of color" is a terrible obfuscation and very inept (especially when compared to a nordic athlete, flushed red after a work out).

But my attempt at this choice of identity failed miserably, since the children were constantly told by their classmates, and sometimes by their teachers, that they simply were not black, and must accept being in some other category - that dreadful word "brown" was offered as a choice. We resisted that, but the problem was never resolved to the satisfaction of all (that, I suppose, is a victory of a sort). So we may not be able to choose some identity unproblematically even when a good reason exists for that choice (in my case, I have to confess, the reason was more political than cultural or morphological), and the limitations of our field of choice may become clearer through practice.

However, all this is not in any way in tension with the basic understanding that we do have choices to make (even though we cannot choose whatever we would like). And the exercise of reasoning and volition can be a very important part of social living, for a life of freedom and, with a little effort, for greater harmony, justice, and peace.

Hills:

The so-called West has never had a corner on serious scientific inquiry or serious technological innovation. And, examples you've presented in The Argumentative Indian *and elsewhere show that it has likewise never had a corner on theoretically-explicit, practically-oriented debate about the core themes of* political philosophy*: religious toleration, personal liberty, the importance of public reason, the answerability of rulers to those they rule, and so on. Those of us who teach political philosophy in Western universities need to appreciate the fact that its central themes have a long and distinguished history outside the West. If we don't, the West may continue to view the rest with a condescension that is theoretically unwarranted and ever more dangerous politically. But this is easier said than* *done. Could any of the texts you've brought to our attention lend themselves to being part of an introduction to political philosophy, political economy, or general ethical theory? If so, how? Do you have any other advice to offer those in search of less ethnocentric ways of teaching these important subjects?*

Sen:

These are very interesting questions. It is indeed very hard to find full texts that can be readily used in a school or college here in the same way as more standard Western texts are used. There are at least three reasons for this difference. First, classical Western terms communicate in a way that imported terms from a distant culture may not. For example, you do not have to explain what "Achilles's heel" refers to, though "Arjuna's eyes" would not communicate much, given what knowledge is standard and what is not in a particular society. (Arjuna, a great prince and warrior in the epic *Mahabharata*, won a major archery contest without actually shooting an arrow when he answered the question put to all contestants about what they could see when they were aiming at the target: he was the only one to say that his eyes could see nothing whatsoever other than the target.) So communicability is an important issue.

Second, the ancient Greeks may have some competitors in many fields in which they shined, but there are some areas in which they were not only totally unusual, but influenced the entire tradition of schooling in Europe and the West, often through the Romans. I do not know of any one who compares with Herodotus, the Greek historian born in Halicarnassus in the fifth century BC, for establishing the discipline of history and influencing the writing of history over two and half thousand years. Similarly, Aristotle was not only an extraordinary thinker, but his way of writing usable treatises, like *Nicomachean Ethics* or *Politics*, surely did influence the stylistic side of presentation of texts as well. The fact that the West certainly has a more well developed tradition of doing texts for pedagogy is both important to recognise (the cultures are not all "similar" in all respects), and has to be distinguished from the presumption that since Confucius or Kautilya did not write in the same way, their ideas must be less important.

Third, we live in a world in which the last two or three hundred years have produced, through the European Enlightenment, industrial revolution and educational transformation, a massive scholarship of disciplined classification and integration linking the Western classics with contemporary concerns. However, if one were living in another time, the picture could be very different. For example, there was something quite unique when Madhavacarya in fourteenth century India wrote his *Sarvadarsanasamgraha* ("Collection of All Philosophies"), with fourteen chapters devoted respectively to fourteen different views on religious philosophy and epistemology (each chapter defensive of the school of thought championed in that chapter, to be criticized in other chapters). The first chapter of the book, which is on agnosticism and atheism, is one of the finest - and deeply sympathetic - presentation of the main arguments in favor of these positions. The second chapter on Buddhism is a similarly sympathetic and efficient presentation of its lines of arguments. And so on through the chapters. Written by a practicing Hindu of the Vaishnava school of thought, the scholarly detachment in presenting the arguments in these diverse chapters remains hugely admirable even today, though the world has moved on in the last six hundred years. Of course, Madhava's own reading of non-Indic-origin philosophies was quite limited - I am commenting not on the width of coverage, but the fairness and efficacy of what the book did cover. I have used extracts from the book in my teaching on modern epistemology, as I have also used extracts from the epistemological writings of the second century philosopher, Nagarjuna (not to be confused with the chemist also called Nagarjuna in the eighth century, whose writings were so influential in the development of chemistry in the Arab world in the following centuries).

So the problems referred to in these questions are real, and the practical implications are also important. We have to make much greater use of substantial extracts from various writings of originality and reach in non-Western literatures, in addition to using good secondary books on the subject (there are, by the way, some very good specialist treatises - one that has just come out and for which I have been privileged to write the Foreword is Bruce Rich's *To Uphold the World: The Message of Ashoka and Kautilya for the 21st Century*, just published by Penguin/Viking). Using extracts is now standard enough in many Western universities when it comes to the dominant religions of the world, but since this is not much done in any field outside the standard domain of sanctified religions, the asymmetric focus on religion alone gives the students—and scholars generally—a very biassed view of the domain and stretch of non-Western intellectual traditions.

Hills:

In Section 5 of your Nobel Lecture you describe the general methodology of formal social choice theory in striking and (to me) some what puzzling terms:

> The general relationship between possibility and impossibility results also deserves some attention, in order to understand the nature and role of impossibility theorems. When a set of axioms regarding social choice can all be simultaneously satisfied, there may be several possible procedures that work, among which we have to choose. In order to choose between the different possibilities through the use of discriminating axioms, we have to introduce *further* axioms, until only one possible procedure remains. This is something of an exercise in brinksmanship. We have to go on cutting down alternative possibilities, moving—implicitly—*towards* an impossibility, but then stop just before all possibilities are eliminated, to wit, when one and only one option remains.[1]

The mathematician in me can readily see why one might appreciate a problem that admits only one correct answer. But the politician in me is tempted to say that practically urgent social and political problems often have many equally good solutions—all of them, alas, unattainable in practice. (If two of them were fully attainable, it wouldn't matter which was in fact attained; we could happily flip a coin.) Could you tell me a little more about how this brinksmanship works in practice, and why it can be expected to yield results of more than mathematical interest?

Sen:

First, a point of clarification. In the passage quoted by you, I am not talking about the desirability of insisting that a problem should admit "only one correct answer" (on that hugely important issue, more presently). This passage, rather, is about how to get "a full axiomatic determination of a particular method of making social choice" (this explanation is on the same page, page 74, from which the quotation is taken). For that specific purpose, that is for the axiomatic derivation of one particular rule (excluding all others), it would be necessary to get a set of axioms that is met by one—and only one—rule (not more than one, since

1 Amartya Sen "The Possibility of Social Choice," in *Rationality and Freedom* (Cambridge: Harvard, 2002), 74.

this will not yield an axiomatic determination of one specific rule). This is just a mathematical point about the nature of axiomatic derivation, not about whether we should invariably seek one—and only one—acceptable answer to a social or political problem.

Coming now to the substantive issue, I entirely agree that in many social or political problems our reasoning, while important in cutting out some proposed solutions, would often enough leave more than one solution in the acceptable category. Indeed, this plurality is an important point championed in my forthcoming book *The Idea of Justice*. (In order not to confuse anybody, I should explain that this book has been announced under various titles over the last decade, including *Freedom and Justice*, and more recently as *Reasons of Justice*—I have been talked out of that name when it was explained to me that it would deeply frighten—and put off—non-specialist readers, which seems sad to me!) The recognition that a complete theory of justice may yield an incomplete ranking of possible alternatives is one of the central concerns of this book (oddly enough—or perhaps not so oddly—this was also one of my major concerns in my first book on social choice: *Collective Choice and Social Welfare*, 1970).

This is, in fact, one of the main points of difference between my approach to the theory of justice and the mainstream theories of justice, such as those of John Rawls, or Robert Nozick, or Ronald Dworkin, or Thomas Nagel. The mainstream search has been principally for one specific set of principles of justice that would univocally determine a structure of just institutions for a just society. The need for such a specific system to be put in place by a sovereign state has been a central feature of mainstream theories of justice at least from Hobbes onwards. And it is the main reason why even as much of a globalist as Thomas Nagel argues "the idea of global justice without a world government is a chimera." So your question in favor of what I call "plurality of impartial reasons" (that is the title of one of the chapters of my new book) has huge implications not only for what kind of answer to seek from a theory of justice, but also for the possibility of advancing global justice even in the absence of a global government. I have discussed this question in an essay called "What Do We Want from a Theory of Justice?" *Journal of Philosophy*, 103, May 2006.

Hills:

One strand of your theoretical work concentrates on the role fundamental human capabilities *ought to play in deciding (a) which cases of* material inequality *constitute relative deprivation of a morally suspect kind, and (b) which cases of* relative deprivation *constitute violations of* fundamental human rights*. One strand in your empirical work, the work on famine, concentrates on the role political arrangements that violate already recognized human rights appear to have played in permitting or fostering the most disastrous kinds of relative deprivation. Could you comment briefly on how you view the methodological relations between these two aspects of your own work as an economist? In particular, how large a role should the political economy of catastrophic conditions like war and famine play in efforts to refine and extend systematic account of what we humans have coming to us as a matter of right? How do these aspects of your work speak to thinkers who take the politics and political economy of catastrophe very seriously, yet find the very idea of basic human right conceptually suspect? (I'm thinking here of Raymond Geuss, but there may well be others.)*

Sen:

Catastrophic situations are important for theory for many different reasons. But certainly one reason for their importance is that even when a theory leaves room for disagreement on some matters (as discussed in the last answer), there could be a general agreement—between competing principles and also between champions of different approaches—that something specific must be done right away. For this reason, even rather complex theories can yield simple conclusions when faced with the extreme problems of a big disaster. Judgments can be a lot more complicated when there are small conflicting arguments that go in different directions, but none of which are strong enough to overshadow competing concerns.

You mention that even those who favor catastrophe prevention very seriously may find "the very idea of basic human rights conceptually suspect." But since you don't specify why they find the idea so suspect, I am not sure I can easily answer your question about how I should respond to this. Your pointer to Raymond Geuss gives me some clue about what you are thinking, but I am not going to try here first to articulate the difficulties that Geuss finds in the idea of human rights, and then discuss how these difficulties may be addressed—that would be quite a long haul (since Geuss, as a powerful thinker, presents complex arguments, rather than simple slogans!). But I would point to two of my essays where I have discussed the various objections raised about the acceptability of the idea of human rights. They are: "Elements of a Theory of Human Rights," *Philosophy and Public Affairs*, 32 (2004), and "Human Rights and the Limits of the Law, *Cardozo Law Review*, 27 (April 2006).

By the way, since you refer to the concept of capability, which - I have argued - is important for some problems, I should mention that many of the serious claims to human rights are not about capabilities at all, but about the rightness of the processes involved (this is *inter alia* discussed in the first of the two papers cited earlier). On a related, but more general, point, while I have been closely linked with the recent developments on work on capabilities (and judge these works to be, in general, very good and illuminating), my theory of justice is not just a capability-oriented theory (many of the other elements—and other differences from the mainstream—are at least as important).

Rob Reich:

Ernest Gellner once wrote, "If the several thousands or more of professional philosophers in America were all assembled in one place, and a small nuclear device were detonated over it, American society would remain totally unaffected."[2] *Agree or not? If true, is professional philosophy (or professional economics) the worse for it?*

Sen:

This is an easy question to answer. Ernest Gellner was a good friend of mine, but I shall have to tax my friendship too much if I had to defend the view that Ernest ever tried very hard to find out what philosophy is about. So, no, I do not agree, and my denial is not only about Ernest's claim that with such precision-bombing of philosophers (but, I take it, no sociologists) "American society would remain totally unaffected." Clarifying what people argue about and how they can understand the problems—and each other—better is certainly an extremely important job, in which philosophers are engaged with some—if varying—success, and to deny the importance of all this would require a level of obduracy or

2 Ernest Gellner, *The Devil in Modern Philosophy* (Boston: Routledge, Kegan, and Paul, 1974), 37-38.

blindness that would be hard to imagine. But you must also bear in mind that my friend Ernest Gellner often made remarks to get an energetic discussion started —I was not against this dialogic device, many of which have taken place in my living room (on one occasion, in my living room in Cambridge, Mass, some of the Junior Fellows of the Harvard Society of Fellows were so driven by Ernest's triumphant dismissals—Kant, Marx, Shakespeare—that I turned pessimistic about whether the evening held any promise of ending).

Dualist Staff:

Today's immigrant societies make multiculturalism a fact of modern life. You have suggested we replace simple group identity with plural affiliation, in which religious, professional, and community-based affiliations all define individual identity. What role does national identity have to play in this picture? What value does nationalism continue to hold?

Sen:

Nationalism is both a curse and a boon. There are distinct ways in which the two different types of effects of nationalism may work. Our national identity is one of the many identities that we have, and nationalism operates mainly through giving special priority to our national identity over other demands on our affiliative attention. Nationalism would tend to be least productive—indeed thoroughly counterproductive—when the main confrontations are along the lines of national divisions themselves (as was the case, for example, in Europe during the First World War), since greater nationalism would only add fuel to fire. On the other hand, nationalism can be productive enough in many contexts, especially when the social divisions and hostilities, within a country or across the world, tend to be based on *other* identities, such as religion or community or ethnicity (as is, to a great extent, the situation right now). The curse and the boon are, in this sense, two sides of the same coin, and depending on the circumstances involved, they can have strongly negative or hugely positive effects.

To understand the reasons for the difference, we have to ask: does the focus on national identity enrich the multiple identities of a person and add dimensions to the picture that would be otherwise neglected? This can be important not only for self-understanding, but also for actual politics. For example, India would not have been able to emerge from the last general elections, in 2004, with a Muslim President, a Sikh Prime Minister and a Christian leader of the ruling party (none of them belonging to the majority community in India of more than 80 per cent Hindus) but for the constructive effects of nationalism (the top office holders were all seen as Indians, rather than as just members of their respective minority communities). On the other hand, to the extent that the India-Pakistan confrontation is worsened by nationalistic thoughts, a weakening of belligerent nationalism would make a good positive contribution in reducing that confrontation. More nationalism during 1914-18 would have been a terrible addition.

We have reason to resist the tendency, common in some circles, of seeing nationalism as an unmitigated evil, and also the tendency, prevalent in other circles, of considering nationalism to be a universal virtue. Nationalism can be mainly a boon, or mainly a curse, depending on the actual circumstances. The central question is whether the development of nationalistic thoughts would sharpen divisions around national identities, or reduce the hold of divisions based on other identities which could be ameliorated by national unity.

RESOURCES FOR PHILOSOPHY UNDERGRADUATES

This section includes listings of journals, contests, and conferences—all of *which are available to undergraduates in philosophy. If you have comments, suggestions, or questions, or if you would like to be listed here in the next issue, please contact us and we will gladly accommodate your request.*

JOURNALS:

There are numerous journals, published both in print and online. The information is as recent as possible, but contact the specific journal to ensure accurate information.

Aporia: Brigham Young University. Submissions due early fall. Papers not to exceed 5,000 words. Send submissions to: Aporia, Department of Philosophy, JKHB 3196, Brigham Young University, Provo, UT 84602. Visit: http://aporia.byu.edu/

The Bertrand Russell Society Quarterly: Edinboro University of Pennsylvania. Visit: http://www.lehman.edu/deanhum/philosophy/BRSQ/

Cyberphilosophy Journal: University College of the Cariboo. No posted due date. Accepts articles, books, web sites, etc. See http://www.cariboo.bc.ca/cpj/.

The Dialectic: University of New Hampshire. Submissions due in April. Essays (15-20 pages), short critical articles, book reviews, artwork. Send submissions to: The Dialectic, c/o Department of Philosophy, University of New Hampshire, Hamilton Smith 23, Durham, NH 03824. Visit: http://www.unh.edu/philosophy/Programs/dialectic.htm

Dialogue: Phi Sigma Tau (international society for philosophy). Published twice yearly. Accepts undergraduate and graduate submissions. Contact a local chapter of Phi Sigma Tau for details or write to Thomas L. Predergast, Editor, Dialogue, Department of Philosophy, Marquette University, Milwaukee WI 53233-2289. Visit: http://www.achsnatl.org/society.asp?society=pst

Discourse: University of San Francisco. *Discourse* is an interdisciplinary philosophy journal featuring the work of undergraduate and graduate students. Send questions to discourse@usfca.edu. Mail submissions to: Discourse, Department of Philosophy, University of San Francisco, 2130 Fulton Street, San Francisco, CA 94117. See also http://www.usfca.edu/philosophy/discourse/index.html.

The Dualist: Stanford University. Submissions due early 2009. 10-30 page submissions. For more information, see http://www.stanford.edu/group/dualist/ or contact the.dualist@gmail.com. Check website for information on submitting a paper and updates on the submission deadline.

Ephemeris: Union College. For more information, write: The Editors, Ephemeris, Department of Philosophy, Union College, Schenectady, NY 12308. Visit: http://www.vu.union.edu/~ephemeris/.

Episteme: Denison University. Due November 14. Maximum 4,000 words. Contact: The Editor, *Episteme*, Department of Philosophy, Denison University, Granville, Ohio 43023. Visit: http://www.denison.edu/philosophy/episteme.html

Ergon: University of South Carolina. Submissions accepted from undergraduate and graduate students. Visit: http://www.cla.sc.edu/PHIL/ergon/index.html.

Interlocutor: University of the South, Sewanee. Direct questions to Professor James Peterman at jpeterma@sewanee.edu. Send submissions to Professor James Peterman, Philosophy Department, 735 University Avenue, Sewanee, TN 37383-1000. Visit: http://www.sewanee.edu/Philosophy/Journal/2006/current.html

Janua Sophia: Edinboro University of Pennsylvania. Submissions and inquiries sent to Janua Sophia, c/o Dr. Corbin Fowler, Philosophy Department, Edinboro University of Pennsylvania, Edinboro, PA 16444. Visit: http://www.edinboro.edu/cwis/philos/januasophia.html

Meteorite: University of Michigan, Ann Arbor. Submissions due February 1. Direct questions to editors@meteorite.com and send submissions to Department of Philosophy, C/O Meteorite, University of Michigan, 435 South State Street, Ann Arbor, MI 48109-1003. Visit: http://www.meteoritejournal.com/.

Populas: University of California's Undergraduate Journal. Submissions limited to University of California students. Visit: http://philosophy.ucdavis.edu/ugjournal.

Princeton Journal of Bioethics: Princeton University. Visit http://www.princeton.edu/~bioethic/journal/.

Prolegomena: University of British Columbia. Visit http://www.philosophy.ubc.ca/prolegom/ or write prolegom@hotmail.com or Prolegomena, Department of Philosophy, 1866 Main Mall, Buchanan E370, University of British Columbia, Vancouver, B.C., Canada. V6T 1Z1.

Promethius: Johns Hopkins University. *Prometheus* strives to promote both undergraduate education and research, and looks for submissions that originate from any scholarly field, as long as those submissions clearly demonstrate their applicability to philosophy. Visit http://www.jhu.edu/prometheus/. Write prometheusjhu@hotmail.com or Prometheus, c/o Philosophy

Dept., 347 Gilman Hall, Johns Hopkins University, Baltimore, MD 21218.

Sophia: University of Victoria. Visit: http://web.uvic.ca/philosophy/sophia/

Stoa: Santa Barbara City College. For more information, write The Center for Philosophical Education, Santa Barbara City College, Department of Philosophy, 721 Cliff Drive, Santa Barbara, CA 93109-2394. Visit: http://www.sbcc.edu/philosophy/website/CPE.html.

The Yale Philosophy Revew: Submissions due February 14. Visit: http://www.yale.edu/ypr/submission_guidelines.htm

CONFERENCES:

There are many undergraduate conferences, so contacting the philosophy departments of a few major schools in a particular area or researching on the web can be quite effective. The conferences below are by no means an exhaustive list.

American Philosophical Association: The APA website, http://www.apa.udel.edu/apa/opportunities/conferences/, contains an extensive list of conferences.

Butler Undergraduate Research Conference: Butler University. Conference in mid-April. See http://www.butler.edu/urc/index.html for details.

National Undergraduate Bioethics Conference: Notre Dame. Visit http://ethicscenter.nd.edu/events/nubec.shtml or write bioethic@nd.edu.

Ohio University Student Conference on Applied Ethics: Ohio University. Conference in late April. Contact ethics@ohio.edu. Visit http://freud.citl.ohiou.edu/ethics/conferences.php

Pacific University Undergraduate Philosophy Conference: Pacific University. Conference in early April. Visit http://www.pacificu.edu/as/philosophy/conference/index.cfm for details.

Rocky Mountain Philosophy Conference: University of Colorado at Boulder. Visit: http://www.colorado.edu/philosophy/rmpc/rmpc.html

Stanford Undergraduate Philosophy Conference: Stanford University. Conference date to be decided, possibly in the late winter or early spring of the 2008-2009 academic year. Contact the.dualist@gmail.com and check the website for more information: http://www.stanford.edu/group/dualist/

ESSAY CONTESTS:

The essay contest listed below aims at a broad range of undergraduates, but there are many other contests open to students enrolled at specific universities or interested in particular organizations. .

Elie Wiesel Essay Contest. Open to undergraduate juniors/seniors with faculty sponsor. Questions focus on current ethical issues. Due in late January. Top prize $5,000. For more information, visit: http://www.eliewieselfoundation.org/EthicsPrize/index.html

The Dualist would like to thank the following contributors from Stanford University:

Philosophy Department
The ASSU Publications Board

Special Thanks:

Vera Haugh

Nadeem Hussain

Natasha Lee

Eve Scott

Kenneth Taylor

Sunny Toy

THE DUALIST is a publication dedicated to recognizing valuable undergraduate contributions in philosophy and to providing a medium for undergraduate discourse on topics of philosophical interest. It was created by students at Stanford University in 1992 and has since featured submissions from undergraduates across North America. If you would like to receive an issue of *THE DUALIST* or to submit a paper, please contact us at the address below. We prefer that submissions be formated acccording to the *Chicago Manual of Style* guidelines. See http://www.chicagomanualofstyle.org/ Papers should be submitted in electronic form only.

Visit our website for submission information:
http://www.stanford.edu/group/dualist

Please email us with any inquiries:
the.dualist@gmail.com

Or write to:

The Dualist
Philosophy Department
Stanford University
Stanford, CA 94305