FORMAL PHILOSOPHY

edited by

Vincent F. Hendricks

John Symons

Automatic Press ♦ VIP

Automatic Press ♦ $\frac{V}{I}$P

Information on this title: www.formalphilosophy.com

First published 2005

Printed in the United States of America
and the United Kingdom

ISBN-10 87-991013-1-9 hardback
ISBN-10 87-991013-0-0 paperback

Typeset in $\LaTeX 2_\varepsilon$
Cover photo and graphic design by Vincent F. Hendricks

Contents

Preface

In the spring of 2005 we had the opportunity to work collaboratively on problems related to the application of epistemic logic and elements from formal learning theory to traditional epistemological questions. Given the nature of this topic, our conversations regularly turned to the more general question of the relationship between formal methods and philosophical investigation. We realized that some of the philosophers we most admire had never explicitly articulated their views on these questions and it occurred to us that it might be worth asking them. We decided to pose five relatively open and broad questions to some of the best philosophers who make formal methods a centerpiece in their work. This book contains their responses to our questions.

The book is motivated by our curiosity but also by our discontent. Neither of us is content with the prominent histories of analytic philosophy currently on the market and we both believe that the discussion of general methodology of philosophy is in a pretty poor state. One of the most significant faults we see with such recent work is its failure to recognize and tackle the central place of formal methods. Shopworn narratives about the failures of logical positivism, the decline of formal methods in philosophy and the rise of intuitions-based conceptual analysis, are neither entirely true nor particularly helpful. In any case, such talk has been overwhelmed by the ongoing buzz of interesting work from philosophers who look much more like Russell and Carnap than Rorty. We hope that this project can serve as a counterweight to some of the more popular surveys of the philosophical landscape. However, our intention is not to promote the use of formal methods in philosophy. Firstly, it is not necessary for us to do so. Formal philosophy is thriving without any advertising. In our view, rather than promoting this kind of work, we can help to begin a fruitful conversation about the deep and interesting methodological problems that formal work in philosophy presents.

19

Wolfgang Spohn

Professor for Philosophy and Philosophy of Science

Department of Philosophy

University of Konstanz, Germany

Why were you initially drawn to formal methods?

My occupation with God and religion, quite intense for a child, I recall, was somehow exhausted at the age of 14, forever, I suppose. My interest then turned to philosophy, a most natural follow-up. It was vacillating at first and without much guidance. With 17 I read the *Hauptströmungen der Gegenwartsphilosophie* by Wolfgang Stegmüller (1952). He presented a number of philosophers, Husserl, Scheler, Heidegger and others, which I found quite obscure. Chapter IX, though, was devoted to Rudolf Carnap and the Vienna Circle. I think I was imprinted by this chapter like the little duck is by the call of its mother. I was firmly decided since to study philosophy at Stegmüller's institute. 38 years later and after rich reflection and experience I still cannot think of a sounder way of doing philosophy than I encountered there.

What example(s) from your work illustrates the role formal methods can play in philosophy?

I am firmly convinced of the crucial importance of formal methods for philosophy. I know many excellent philosophers who do not use formal methods and are not able to profitably do so, and I know a lot of formal papers that are terribly boring, since they neither make much philosophical sense nor are particularly deep from a formal point of view. Still, properly applied formal methods achieve an inestimable philosophical surplus out of reach of any other means. Of course, different philosophical fields are amenable to formal methods to largely varying degrees, and the most difficult and valuable work often consists in making a field so amenable in the first place. Generally, I believe that the fruitfulness of formal methods extends much further than is usually thought. Clearly, though, one needs a good sense of self-constraint.

As far as my own papers are concerned, I sense that I tend to overdo the formal methods (to the detriment of their accessibility). There are great masters of leaving formal methods implicit; compare, e.g., the formally explicit book of Lewis (1969) and his formally implicit twin paper (1975). However, there is a fine line between leaving formal methods implicit and not having them in the back at all, and I want to keep a clear distance from this line by all means. Whence my tendency.

Own examples? The example I have made the biggest fuss about is ranking theory. This is an elaborated account of Baconian probability (as opposed to the real Pascalian probability) or theory of the rational dynamics of (plain) belief. Since any theory of doxastic change is nothing but an account of the problem of induction, the importance of this example is obvious. However, let me refer to my most recent survey article Spohn (forthcoming *b*) concerning that example and turn to another less known one.

It is about decision theory. Since Nozick's (1969) famous presentation of Newcomb's problem philosophical decision theory is split between causal decision theory (the majority opinion) the adherents of which recommend two-boxing and evidential decision theory some adherents of which favor one-boxing. The issue is still contested, mainly, I think, because the majority is not stable, but rather plagued by the question "Why ain'cha rich?" which is not soothed by the answer: "Because irrationality is rewarded in this situation." In Spohn (2003) I have developed a sophisticated argument *justifying one-boxing within causal decision theory*. Here, I want to give a very rough sketch of the argument in order to afterwards explain the quite obvious role of formal methods in this example. The sketch proceeds in eight steps:

The *first* ingredient of the argument is the theory of *Bayesian nets*. A Bayesian net is a directed acyclic graph the nodes of which represent factors or (random) variables plus a probability measure over the algebra of propositions (or events) generated by these variables such that the measure agrees with the graph. The latter means that the set of parents of each node is the smallest set such that this node is probabilistically independent from all its non-descendents given this set; cf. Pearl (1988, sect. 3.3). The fact that this notion is well defined and that indeed for each measure there is such a graph is due to the so-called graphoid axioms for conditional probabilistic independence among sets of variables; cf. Pearl (1988, sect. 3.1). These properties are crucial for the whole mathematics of Bayesian nets and were already discovered by Dawid (1979) and Spohn (1978, 1980).

The *second* step is that Bayesian nets have a *causal interpretation* provided that the nodes or variables are temporarily ordered and the vertices agree with the temporal order, i.e., the endpoint of a vertex is always later than its starting point. The causal interpretation then is simply that each vertex represents a direct causal dependency; its endpoint directly causally depends on the starting point. At this point, however, I have an argument with most other causal theorists working in this paradigm. The common opinion is that a vertex in a Bayesian net is symptomatic of a causal dependence (cf. Spirtes et al. 1993, ch. 3, and Pearl 2000, ch. 1), whereas I say that it is definitive of a direct causal dependence. (I have more carefully analysed this difference in Spohn 2001)

My stronger claim makes sense only because, in the first analysis, I take causal dependence to be frame- or model-relative; insofar only the variables in the Bayesian net are considered and no more, the causal dependencies run as represented in the net. Of course, we think that causal dependence is an objective relation in the world not relative to the frame we happen to consider. The only way I can do justice to this thought is by referring to the universal frame embracing all variables whatsoever. The universal Bayesian net, as it were, displays the causal dependencies as they objectively are.

One may wonder whether the universal frame is really well defined. Therefore, I am, in a *third* step, more interested in how the causal dependencies relative to smaller and larger frames relate. This can only be judged from the point of view of the larger frame. Hence, I am more specifically interested in the *reduction of Baysian nets*. That is, if one node of a Bayesian net is deleted, how do the vertices get rearranged? Basically, the answer is: the vertices between the remaining variables stay the same, and if C is the deleted node, then $A \to C \to B$ reduces to $A \to B$ and $A \leftarrow C \to B$ reduces to $A \to B$ (provided A precedes B). Reversely, this means that, whenever there is an apparent direct causal dependence $A \to B$ relative to a small frame, this may turn into an indirect causal dependence $A \to C \to B$ or into a common cause relation $A \leftarrow C \to B$ relative to an enlarged frame.

Well, this is roughly so. There is a third possibility that turns out to play no role for the rest of my argument. And there are conditions to my claims that should be carefully considered.

Now, what have Bayesian nets and their causal interpretation to do with decision theory? This is the *fourth* consideration. Let some Bayesian net be given and suppose that you can directly manipulate some variable. You can choose any value for it you like. We might then call that variable an action variable. What should you do? This depends, of course, on your utilities for all possible realizations of the variables in the net. And it depends on your beliefs. Should you just use the probabilities conditional on the envisaged value of the action variable in order to determine the conditional expected utility of that value? And maximize expected utility on this basis?

In opposition to evidential decision theorists, the causal decision theorist says no, you must not. By manipulating the action variable directly, you cut it off from its causal dependencies as represented in the Bayesian net, you treat it as uncaused by the other variables in the net or exogenous and at best caused only by you or your will. This means, however, that instead of the original Bayesian net you have to consider the so-called *truncated Bayesian net* which one obtains from the original one by deleting all vertices ending at the action variable and by substituting the original probability measure by its so-called truncated factorization that agrees with the truncated graph. The terminology is that of Pearl (2000, sect. 3.2) where the procedure is described in detail, but the substance may already be found in Spohn (1978, sect. 5.2). It should be clear that it is the reasonability of just this step that is at issue between causal and evidential decision theory. And I clearly take the causal side.

So far, we have only prepared the grounds for the argument to come. My point will be that the combination of reduction and truncation produces most interesting effects. So, let us, in a *fifth* step, start from the decision theoretic point of view, i.e., from a truncated Bayesian net with exogenous action variables. And let us think about how one might undo the truncation. This means thinking about on what the action variables causally depend. At first, this appears irrelevant from the decision theoretic point of view. When taking a decision, one does not think about the possible causes of one's actions, one rather evaluates the possible

actions (according to their expected utility) and decides for one of the best. Still, the consideration will bear fruits:

In a way, it is obvious how our actions are caused. An action is caused precisely by the decision situation in which it is best, where a decision situation is a subjective state of the agent consisting of desires and beliefs and represented by a truncated Bayesian net with a distinguished action variable and a utility function. Hence, it is also clear how to complete the truncated Bayesian net. Each action node is to be preceded by a decision node, a variable taking as values all possible decision situations deciding about that action node. It is precisely this decision node on which the action node causally depends, indeed directly, as long as we do not attend to how mental states eventually issue in bodily movements.

I call the truncated Bayesian net (plus a utility function) a *simple decision model* and the completed Bayesian net (plus a utility function) a *reflexive decision model* because it contains nodes reflecting on the truncated simple decision models. Fully spelled out, such reflexive decision models are enormously complex, but most interesting structures which I strongly recommend for further inquiry.

Here, however, we need not consider the whole of that complex structure. Let us, in a *sixth* step, only focus for a while on the *causal place of these new decision nodes*. The following assumptions are natural and crucial.

1. Each action node has exactly one decision node as a parent. It cannot have two decision parents; you cannot decide twice about the same action, the first decision would then be causally idle and not a genuine decision. It must have at least one decision parent, because rational (or irrational) action is precisely characterized by being caused in this way. And it cannot have other than decision parents; decision situations are the complete direct cause of the ensuing action.

2. Each decision node has at least one action node as a child, since a decision must decide about some action. It may have several action children; one may decide about several actions or about a whole course of actions at once.

3. Each decision node may have many parents. Indeed, a decision situation is a very rich item causally depending on a host of other factors; simply consider in which complex ways your beliefs and desires are caused.

4. Each decision node may have other than action children. Your propositional attitudes are not exclusively expressed by your actions. You may look a certain way when you are determined. Some uncontrolled behavior, your mimics, etc., may be indicative of your desires. Other persons may observe all this and draw their conclusions. Or think of the lie detector assessing your beliefs not via your actions. And so on.

5. No decision node need temporally immediately precede its action children; I can now decide what to do tomorrow or in a month or in five years. One must not confuse here causal and temporal immediacy. Of course, when I now decide what to do in five years, decision and action must somehow be mediated, by memory, by staying determined, etc. However, this is only a difference of degree to allegedly immediate decisions, as everyone painfully knows who forgets from one second to the other what he wanted to do. This mediation is not modeled in reflexive decision models.

Each of these assumptions may be contested, and each makes a big difference for the resulting reflexive decision theory.

In a *seventh* step, we must now consider more closely the relation between the truncated simple and the completed reflexive decision model. The simple model is not simply the truncation of the reflexive model. Note that truncation does not diminish the nodes, it diminishes only the vertices and modifies the probability measure so that it agrees with the truncated graph. Yet, the reflexive model, by attending to the causes of actions, additionally contains decision nodes not at issue in the simple model. Hence, we must first reduce the reflexive model by the decision nodes and then truncate it with respect to the action nodes. The simple model is the *truncated reduction* of the reflexive model. The task then is to account for these truncated reductions in detail. I have given a rough idea of reductions and of truncations. However, in combination they generate the final crucial twist that supports my initial claim about Newcomb's problem.

I am not going to explain this final *eighth* step in abstract; my presentation has already become too imperspicuous. Let me rather exemplify it with two important philosophical puzzles, the Toxin case and Newcomb's problem; this at the same time illustrates our abstract considerations so far.

In the *toxin puzzle*, someone, the predictor, approaches you at noon and promises you a lot of money for managing to form the firm intention or to decide before midnight to drink a glass of toxin tomorrow noon, which makes you feel sick for some hours; in any case, the reward by far outweighs the sickness. And, the predictor adds, he has a cute cerebroscope that reliably tells whether you really have the intention. Note the reward is for the right intention, not for the drinking. The puzzle is that you do not seem to be able to get the reward even though you would like it.

Let us represent the situation by a decision model. There are five variables involved: your mental state M before midnight, the signal C of the cerebroscope at midnight, the reward R soon afterwards, your possible drinking T of the toxin tomorrow noon, and your possible sickness S for some time afterwards. My convention is to use squares for action nodes, triangles for decision nodes, and circles for the other nodes. So, prima facie the relevant Bayesian net looks like this:

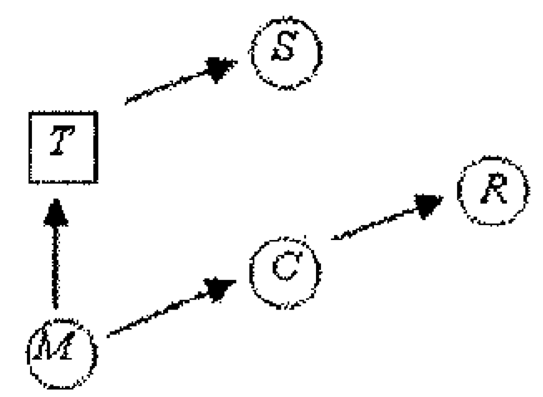


FIGURE 19.1.

However, where is the decision node for the action node T? There are two possibilities. Either, the mental state M is already the intention or decision D to drink, or not to drink, the toxin. Or, as most say, the decision D about T is definitively taken only briefly before tomorrow noon; in that case, though, the mental state M is not one of resolution and the cerebroscope will tell so.

Let me graphically represent the two alternatives together with their truncated reductions (this illustrates at the same time how they work). One model is this: (figure 19.2)

There, in the simple model on the right side it is clear what maximizes conditional expected utility: it is not drinking the toxin, but there is no way to get the reward (unless the cerebroscope makes an error). Hence, those attached to this model despair of getting the preferred reward.

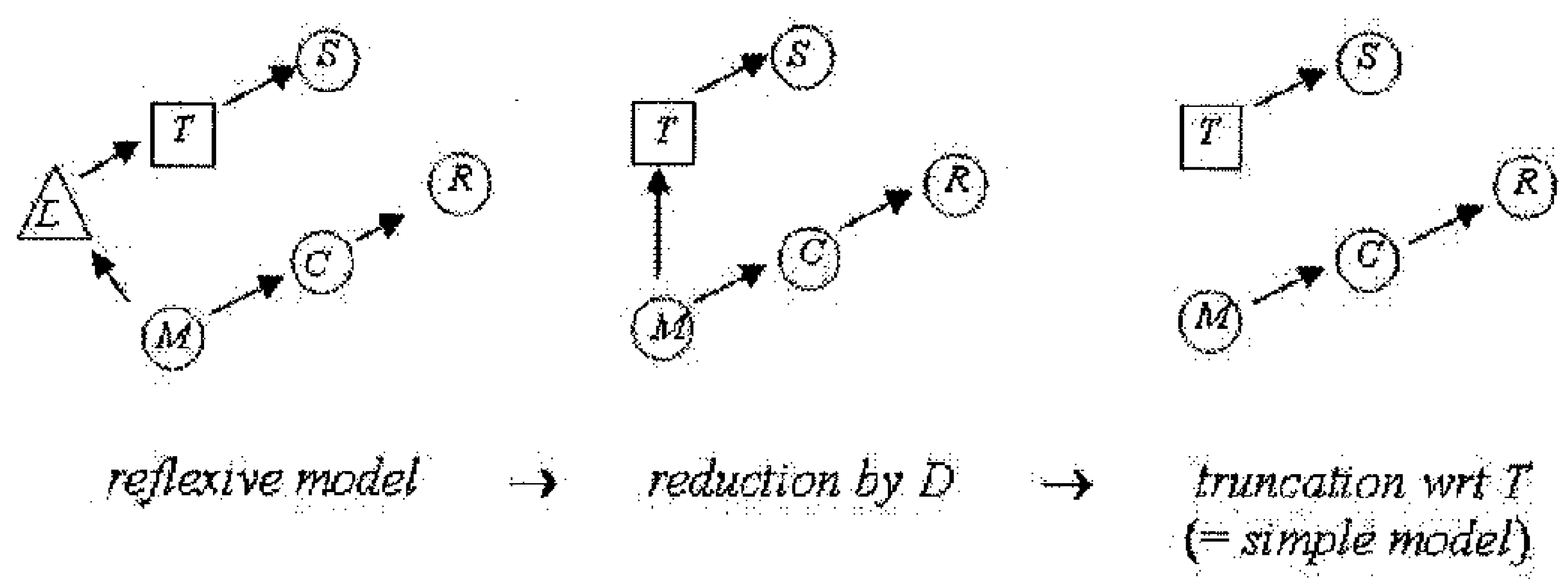


FIGURE 19.2.

The other model is this: (figure 19.3)

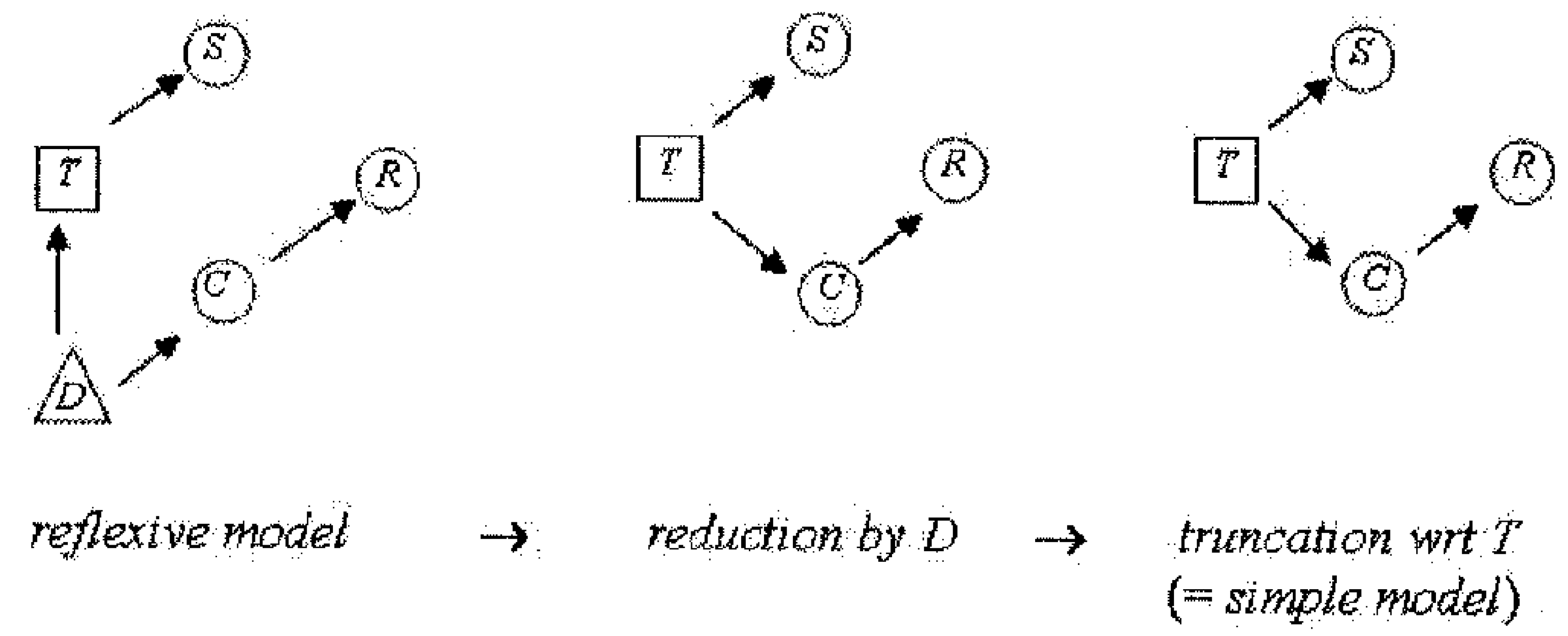


FIGURE 19.3.

We observe here a crucial anomaly in the reduction, a causal arrow from T to C running backwards in time. Do I suddenly plead for backward causation? Of course not. The point is rather this: According to the reduction rules explained above we should have an arrow running from C to T, which, however, would have to be deleted in the truncation. This seems unjustified. Compare this case with another scenario in which the cerebroscope takes 24 hours to process its data and to give a verdict, so that T precedes C. Intuitively, this should make no difference whatsoever. However, the above reduction rules say that reduction results in an arrow from T to C in this scenario, which would not fall victim to the subsequent truncation. Hence, I propose to change the reduction rules *only for this special case* as indicated in the graph above. Recall that an arrow in the reduced graph represents a direct or an indirect causal dependence *or* a common cause relation.

And it represents the latter (and no backwards causation) in the reduced graph. Hence, I am far from assuming causal absurdity.

Once this step is accepted, it is clear that truncation runs empty in this case (since T is exogenous already in the untruncated graph). And it is clear that it is drinking the toxin that maximizes conditional expected utility in the resulting simple model. So, this is what you rationally decide before midnight. By "decide" I mean "decide" without any afterthought or reconsideration, and this includes actually drinking the glass of toxin (or at least seriously attempting to do so).

The crucial point here is the reduction anomaly, the necessity of which emerges only with the subsequent truncation. This is the important special effect announced earlier of combining reduction and truncation.

The story is much the same for *Newcomb's problem*. Its representation by a simple model involves three variables: there is first the prediction P of the predictor and the corresponding filling of the opaque box, a bit later you take the opaque box or both boxes (action node B) and finally you get the reward R depending on the prediction and your action. This yields the simple model:

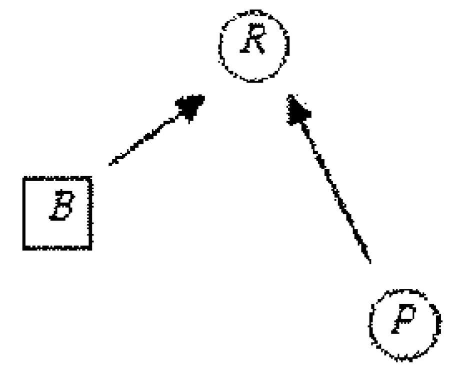
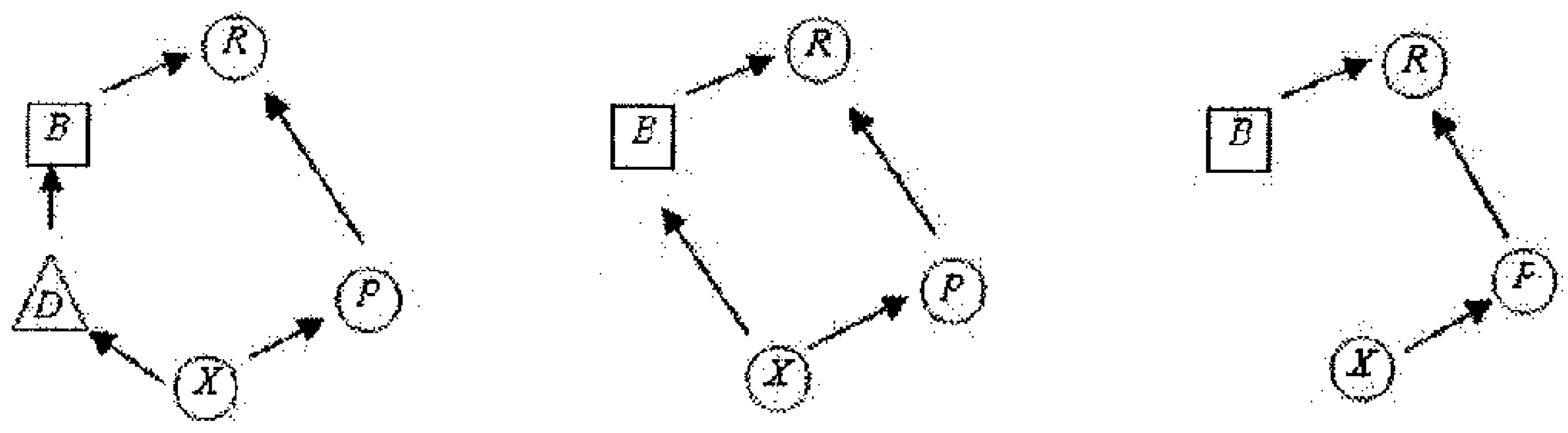


FIGURE 19.4.

in which it is clear that two-boxing maximizes conditional expected utility, as causal decision theorists have said all the time.

How may we extend this simple model to a reflexive one? There are again two ways. The reflexive model must somehow account for the surprising correlation between the predictor's prediction and your action. One possible explanation is to assume some common cause X of the prediction and your decision D. Then the reflexive model and its truncated reduction looks like this: (figure 19.5)

According to the simple model of figure 19.5, as in the model of figure 19.4, only two-boxing is rational. This is precisely the account of Eells (1982, ch. 8). He calls it a justification of two-boxing in terms of evidential decision theory. However, I would rather say that he has thereby laid grounds to reflexive decision

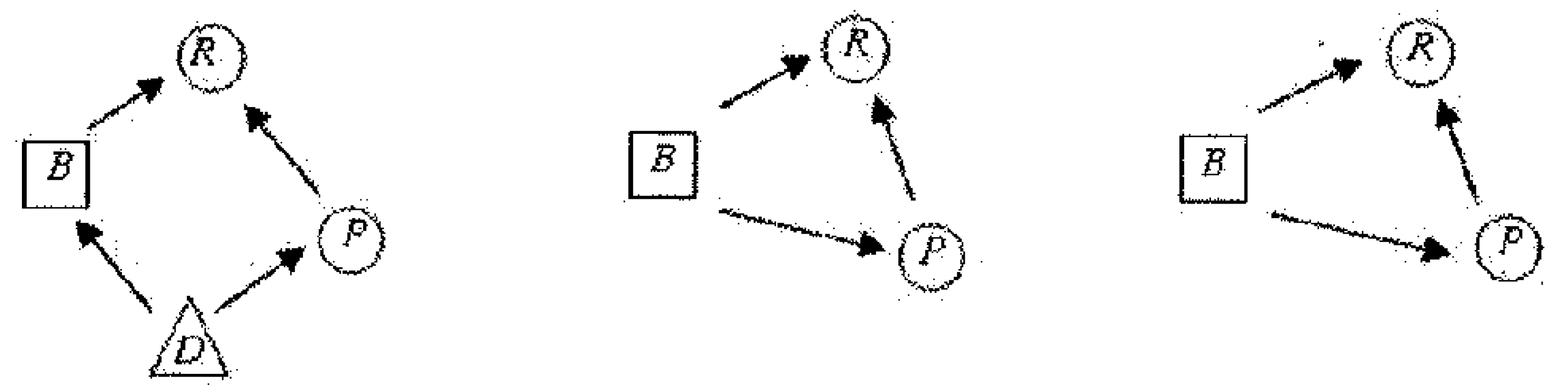


reflexive model → *reduction by D* → *truncation wrt B (= simple model)*

FIGURE 19.5.

theory.

There is another explanation for the correlation, namely that your decision D influences the prediction P. Then the reflexive model and the truncated reduction look like this: (figure 19.6)



reflexive model → *reduction by D* → *truncation wrt B (= simple model)*

FIGURE 19.6.

Note again the reduction anomaly resulting in an arrow from B to P backwards in time, which, however, indicates merely that B and P have a common cause in some decision node. In the resulting simple model one-boxing is obviously the rational thing to do. This is what I take to be a rationalization of one-boxing within (a reflexive extension of) causal decision theory.

You might object that the decision D cannot influence the prediction P because you take it only when you are standing before the boxes. Recall, however, my remarks about the temporal relation between decision and action node. When standing before the boxes you might as well think that as a rational person you were committed all along to take only one box and that the predictor had sufficient time to observe your rationality and to infer your

commitment. I indeed find that this explanation of the correlation between prediction and action is more plausible than Eells' common cause scenario.

The upshot of the whole argument is this: There are rationalizations both of two-boxing and one-boxing according to where you place the decision node. But there is no doubt what the more profitable rationalization is. Hence, since you have a choice where to place the decision node, you should place it so early that you resolve in one-boxing.

Note, by way of comparison, that you cannot thereby rationalize non-smoking in Ronald A. Fisher's smoking gene scenario. In that scenario the smoking gene is a common probabilistic cause of your desire to smoke and the lung cancer. This corresponds to the first reflexive model of Newcomb's problem, and then it is certainly rational to yield to your desire and smoke. If we would want to carry over the second reflexive model, we would have to assume that smoking still does not cause lung cancer, but that the desire to smoke somehow activates the dangerous gene. However, this was not Fisher's story.

So much for my argument. There is no point here in fighting it further (and I am sure it needs further fighting). Let me only add that these considerations have much wider application. For instance, McClennen (1990) has carefully explained two decision rules, sophisticated choice and resolute choice. The latter appeared unintelligible to many commentators. My considerations offer, I think, further rationalization of resolute choice and indeed integrate sophisticated and resolute choice into one model instead of treating them as competitors.

More generally, commitment was tacitly the key notion of my considerations. Commitment is a most central notion to practical philosophy, but it receives only informal treatment and somehow remains ill-understood. My above considerations have the potential, I believe, to improve this situation. Moreover, my real focus in Spohn (2003) was to rationalize cooperation in the single-shot prisoners' dilemma, at least a plausible attempt in view of the fact that the prisoners' dilemma may be conceived as a two-sided Newcomb's problem.

Let us finally return to the original purpose of this exposition. What does it show about the role of formal methods in philosophy? I think, three quite trivial things which I state nevertheless, because they are so clearly exemplified here:

First, it is serious and important philosophical problems that are successfully addressed by formal methods. This is clear from my last remarks; also, Newcomb's problem is not just an odd puzzle, it lies at the heart of our understanding of rational agency.

Secondly, what I have sketched is a relatively sophisticated formal reasoning. By turning it formal, one can and must make explicit all the assumptions and conditions needed. One could certainly give a purely informal version of the reasoning, and this would certainly result in a plausible story. However, an opponent could easily tell an equally plausible counter-story. And the ensuing argument would get hopelessly lost in confusion. I am not claiming that the formal reasoning proves my case. Hardly any philosophical thesis can be strictly proven. Still, the formal reasoning greatly helps in getting clear what the issues are.

The third point is stronger than the second. Even the heuristics of my reasoning requires formal methods. In this case I cannot imagine that I first have an informal sketch of the argument, then try to formalize it, and in the end come up with the formal version. In order to see the argument at all it was important to already have a trustworthy formal model of decision situations in terms of Bayesian nets with action nodes and to think about the formal properties of the model.

In this way, my example shows that formal methods are required and useful at every stage of inquiry.

What is the proper role of philosophy in relation to other disciplines?

Philosophy has many roles and many proper roles, in relation to human life in general, a relation not at issue here, and in relation to other disciplines. Here, I see at least five proper roles:

The *first* important role is that philosophy is a *speculative forerunner* of scientific issues. Thinkers are often beset with pressing questions of a broadly empirical nature, not knowing how to turn them into sound scientific questions to be tackled in a sober empirical way. They engage then in speculation, develop models and perspectives, and thus open possibly quite influential ways of how to think at all about such issues. This is often called philosophy, with some justification, since it requires intellectual freedom (or lack of control) hardly to be found elsewhere.

There are many great historic examples for this role of philosophy, Democritus' atomic theory being perhaps the most famous one. In our times when the disciplines have become extremely

specialized and ramified, this role is certainly diminished. Sketchy philosophical models for our highly advanced and sophisticated cosmology, e.g., would simply be ridiculous. Things differ, though, in relation to psychology. We still find a lot of speculative mental model building within philosophy which is not without influence on psychology.

Philosophy is not only a forerunner, however, it is *secondly* even the *mother* of many scientific disciplines (though I do not claim parthenogenesis, of course). This is a familiar phrase with which one mainly associates the times of Enlightenment, roughly from late 17th to early 19th century, which generated the basic differentiation of scientific disciplines. The philosophical origins of analysis (perhaps the most powerful tool of mathematics), Newtonian physics, psychology, sociology, economics, the political sciences, etc. are well known.

These may seem to be past merits. However, the process continues, perhaps in a less speculative manner, till today; and one should emphasize this point whenever the current social or political significance of philosophy is critically discussed. A prominent case is Artificial Intelligence, which has certainly more than one parent; but one should read how much Marvin Minsky and other fathers of AI say they have learned from philosophy. Another example is linguistic semantics and pragmatics that started establishing as a separate discipline only 30 years ago and are still dominated by ideas developed by philosophers.

Yet, I do not want to deny that times have changed. The current role of philosophy is, *thirdly*, rather that of a *sister* of sciences and humanities. What has been a mere boundary has often developed into a larger research field promoted by philosophy and other disciplines in a cooperative way. Again, there are many important examples. Perhaps the most comprehensive example is cognitive science, a collective enterprise of neuroscience, psychology, linguistics, AI, and philosophy. Science studies are due to the joint efforts of sociology (of science) and history and philosophy of science. Philosophy, economists, and politologists equally contribute to social choice theory. Applied ethics is philosophical ethics in collaboration with medical, biological, environmental, or economic studies, to mention only the more prominent connections. Moreover, there are various issues without a disciplinary name that have turned interdisciplinary. The theory of causation is belabored by philosophers, physicists, statisticians, economists, and even AI researchers. After philosophers have learned general relativity theory, space-time is again a joint topic. And so forth.

It seems obvious to me that philosophy essentially contributes to all these cooperative efforts; they would be poorer without philosophy. This is perhaps an even stronger argument in critical discussions of the current significance of philosophy. However, one must never forget that philosophy does not exhaust itself in mother- and sisterhood. It is able to cultivate these kinships only because of the productivity of its core disciplines: ontology and metaphysics, epistemology, philosophy of mind and language, and ethics. Philosophy stands and falls with its independent worth endowed by these core disciplines.

So far, I have described the kind of relation philosophy has to other disciplines. My examples displayed the legitimacy of these relations. Still, it must be in virtue of some contents that philosophy can engage in these relations, contents delivered by philosophy rather than any of the other disciplines participating. What are they? There are two fundamental kinds of contents, I believe, that characterize the two substantial roles philosophy will forever play for other disciplines.

The first kind of content is *normative content.* Normative discussions take much space in philosophy, and they do so explicitly. Theology and jurisprudence are also firmly aware that they are dealing with normative contents. In all other disciplines, in particular in the natural, but also in the social sciences, the normative dimension remains, as far as I can see, largely implicit; apparently, this dimension does not fit into the picture these disciplines have of themselves. This does not mean, however, that this dimension does not exist. On the contrary, and it is philosophy's task to bring to bear its normative wisdom to the other disciplines (and it is obviously philosophy rather than the other normative disciplines mentioned that is called for here). This, then, is the *fourth* proper role of philosophy: to provide the normative input for the cooperative cognitive enterprise.

In fact, the normative dimension of empirical science has at least three aspects, and at least two are genuinely philosophical:

One kind of normative issue is that of allocation: How much of our bounded means should we spend on the questions in which our unbounded curiosity might take interest? These issues are most involved, often left to the market, often hotly politically disputed. By and large, philosophy has no special competence here. However, such allocation issues often have a moral dimension, I suspect indeed much more often than is usually thought. To this extent

philosophical advice may well be sought.

Another kind of normative issue is that of methodology. Once it is decided which question to investigate, the follow-up question is how to do it: how to set up the inquiry, what to conclude from its possible results, how to assess the various hypotheses at hand, and so on. These are methodological issues, and as such they are normative, though their normative character is usually veiled. For instance, statistics, better manned than philosophy, is fundamentally methodological and hence normative in character, but this fact almost disappears behind all the sophisticated mathematics.

Of course, methodology is no exception to ubiquitous ramification and specialization. However, general methodological issues are genuinely philosophical; they belong to normative epistemology and philosophy of science. And these issues are far from being exhausted. Philosophy has still a lot to teach to, and to learn from, other disciplines about them.

There are, finally, general normative issues: how to behave rationally, how to behave morally or, rather, simply how to behave not only in scientific contexts, but in life as such. Again, on a general level these are genuinely philosophical issues. They may not appear relevant to empirical research (except insofar as researchers are also agents). However, this is not so. As I have argued several times, most recently, albeit briefly in Spohn (forthcoming a), normative theorizing does not only tell how we should behave, but also how we ideally behave; and deviation from the ideal is to be described as such (unless we adhere to an, I think, highly implausible eliminativist program in psychology). In this way, normative theorizing becomes an essential part of empirical research on all human affairs; psychology is essentially normative. This is a third aspect of the normative dimension in which at least the human sciences are tied to philosophy, even though the awareness of this fact remains wanting.

The other kind of content for which philosophy has a special expertise is *modal content.* What do I mean? Well, all scientific disciplines seek the truth, claim to find it, and sometimes actually do; each scholar is a truth authority for his field. The philosopher's task then is to sort out the (alleged) truths found; and this is his authority. There are many kinds of truths, and the modalities are there to classify them. There are necessary and contingent truths, and there are half a dozen different senses of necessity (or more). There are analytic and synthetic truths, and the notion of analyticity is as difficult as the whole theory of linguistic meaning.

There a priori and a posteriori truths, and Kant's suggestion that apriority is wider than analyticity opened new, though debated philosophical spaces. Among the a posteriori truths, one may distinguish empirical and theoretical truths, due to the suggestion that the meaning of theoretical terms differs very much from that of empirical terms. Causation is presumably a modal notion, explanation definitely is. There are methodological, i.e., normative claims which many call true or false as well. And so on. All these distinctions are subsumed under the label 'modality'.

It is difficult to know one's way through the jungle of modalities. Even philosophers who think for decades about them get easily confused. Some claim to have gained clarity in the end; alas, the alleged clarities differ wildly, and so the collective state is again one of confusion. This, however, is due to the complexity of the subject matter, and even though it is contested among philosophers, they know so much more about it than scholars from other fields.

And it is important to accomplish the classificatory work of applying the modal categories. Each modal category has its characteristic grounds of truth, and all scientific discourse essentially depends on the character of these grounds. Arguments for an empirical truth, for instance, radically differ form those for an analytic truth, and both have nothing to do with justifying a normative truth. Hence, discourse is bound to end up in confusion if the modal classification of the focal claims is unclear. Here philosophy is called to bring in its expertise. This is the *fifth* and last proper role I see for philosophy in relation to other disciplines.

Sometimes I sense that scientists feel patronized by this role of philosophy; they suspect that philosophers want to be the better scientists. There are perhaps some such presumptuous philosophers. This would be a severe misunderstanding, though. Philosophy indeed has some competence concerning the claims made in other disciplines, namely in the way indicated. Scientists are well advised to acknowledge this competence, and philosophers are well advised not to confuse this with an authority on the scientific field.

What do you consider the most neglected topics and/or contributions in late 20th century philosophy?

One must be aware how radically the academic condition has changed since World War II. We have seen an unprecedented academic explosion due to increased needs and a fabulous wealth in the western world. I use to say that half of the professional philosophers ever existing are still alive; and though it is difficult to count (who in the ancient or medieval times is a professional philosopher?), my guess is probably not off the mark in magnitude.

Moreover, the communication conditions have changed even more radically. Philosophy has become so easily accessible. The living philosophers have read so much of the dead, I assume much more than the dead read of the dead (whereas the dead had little opportunity to read texts of the living). There are so much more publishers, journals, conferences, guest lectures, etc. Internet and e-mail has further accelerated communication in an unbelievable way.

Often it appears to me that these dramatically changed conditions and relations have received insufficient attention in the still wide spread history-biased understanding of philosophy.

Hence, I find the title question misplaced; it is a question for conditions of scarcity, not for conditions of abundance. And we are living in the latter. A sure sign of this, and one I perceive with great skepticism is the tremendous increase of encyclopedias, handbooks, companions, introductions, etc., in the last 10 or 15 years. If they are well made, one is grateful for them; but they show at the same time that there are many philosophers who have no good idea how to occupy themselves.

What I find much more fascinating is the issue of the power relations in the modern unprecedented philosophical market. From where to where do the influences run? Why do they run as they do? Do the power relations produce systematic distortions or even (unintentional) suppression? Such a market needs and has opinion leaders. How do they get their role? Certainly in virtue of their charisma, their quality und originality. I suspect, though, there are many more factors at work. How, then, do the opinion leaders structure the discussion? For good or for bad? The communication mechanics is presumably not so different in other disciplines. So, which observation can be generalized and which are special to philosophy? These would be the questions to investigate.

I have not seen any study directed to philosophy, though it would be worthwhile. I am certainly not the one to do it. However, I would like at least to mention that one factor appears to me still to be of utmost importance: language. Of course, international communication requires a common language, and as the world has developed (this is part of the power relations), this language is English. There are those who master English perfectly and those who master it imperfectly; the large majority of foreigners belong to the second group. This is an inevitable asymmetry which seems particular relevant for philosophy, because philosophers pay much more attention to phrase and style than many other disciplines and because it is still less natural for philosophers to adapt to a common language than for most other scholars.

The long and the short of all this is that I find it unlikely under the present circumstances that there really are forgotten, though important topics and contributions. Almost the only way how there could be such things is that some philosophers entirely withdraw from the academic fuss and develop their thought in obscurity, which might nevertheless be ingenious. There are such philosophers – should I say: fortunately? – but, of course, it is difficult to know of them.

Actually, I know of at least one: Ulrich Blau from the University of Munich. For almost 30 years he is working on the logic of paradoxes and indeterminacies, pursuing deep perspectives in the theory of truth, semantics, philosophy of mathematics, and much further. To some extent, his ideas are similar to known accounts of semantic paradoxes (though I cannot decide issues of priority), but in many respects they go far beyond. He has published only in German, and the last publications are about 20 years old. Now, finally, his opus magnum, Blau (forthcoming), a book of 1000 pages, the fruit of 30 years of thinking, is intended to go in print. I am unable to reliably assess this incredibly rich work, but I am sure it contains an exceptional lot of ingenuity and definitely deserves much wider attention. This is the first positive answer to the fourth question that comes to my mind.

What are the most important open problems in philosophy and what are the prospects for progress?

I said already that hardly any of the philosophical problems are forgotten or neglected, and I also mentioned that hardly any of the philosophical problems are or will be solved; they will remain open. They are usually not that kind of problem that can be solved; rather, the spectrum of possible answers can be widened, ramified, substantiated, clarified, and united, and this process is clearly a progressive one, even if no one answer can be distinguished as the true one.

Which among all those problems are the most important ones? This is presumably a matter of taste. In any case, I am convinced

that philosophy as a whole will be of increasing importance. I see two historical long-term tendencies in both of which philosophy is centrally involved.

Despite contrary appearances of an increased religious influence on the historic course of events, I am first convinced that Enlightenment is not yet finished and that in the long run secularization will be the dominating tendency (I am a philosopher, after all). My minor reason for this conviction is that growing prosperity and education always has a moderating effect on religious affairs. My major ground, though, is that I think that epistemic rationality is a slow, but very strong force. The major religious doctrines are thoroughly interwoven with epistemic irrationality. This is why I think they are doomed in the long run; we cannot forever maintain beliefs against reason simply for their real or alleged good consequences.

I am not claiming that a secularized world is a better one. I hope that religious mania gets controlled, but I am also opposed to uncontrolled egocentrism on an individual, social, or political level, which often accompanies secularization. I fear, for example, the delusion of eternal life in both forms: as the ultimate ground of many religions conceptions and as an ultimate motive of modern biomedical research (which it surely is).

Here, mankind faces a huge ongoing task. If secularization is unavoidable, at least its bad consequences must be avoided. Philosophy has to certainly play an important role in this task. It is, for instance, often claimed that certain morally valuable attitudes are not to be had without their religious justificatory superstructure. If this were so, all the worse for the attitudes. However, I do not accept the premise; of course, the morally valuable attitudes are amenable to an enlightened justification. And philosophy is there to provide it.

The other great tendency I see at work is that of self-destruction, and I do not know at all which of the two processes, secularization or self-destruction, will be the faster one. It seems obvious to me that mankind has reached a critical stage. There is too much wealth and too much need at the same time, and both are ruining the world. I wonder for how long the earth will carry six billion men (as present) or even nine or ten billion (as predicted); not for many centuries, I fear. This is an even bigger challenge to mankind, to its science and technology, but even more to its social and political forms. Indeed, it will be the biggest challenge we will have ever faced. Mastering it, there is still hope, will require the best of our human capacities, and again philosophy will have to play here an important foundational role.

This sounds as if only practical philosophy would be called upon. However, in the end, theoretical and practical philosophy form an inseparable unity. So, both challenges concern philosophy as a whole. Still, since I am rather a theoretical philosopher and since I guess the title question rather asked for my taste, let me give a second, quite different answer.

Ontology is the one core discipline of theoretical philosophy, epistemology is the other. For me, the deepest problem of theoretical philosophy is the relation of its two core disciplines. The long-known, ongoing wavering between ever more sophisticated forms of realism and idealism is the most obvious symptom of the deepness of the problem. I still think that the rearrangement of ontological and epistemological modalities, of necessity/contingency and apriority/aposteriority, by Kripke (1972) and Putnam (1975) brought revolutionary progress here. By establishing the independence of the two dimensions or distinctions, this rearrangement had a most clarifying and liberating effect; it was a break with centuries old entanglements.

At the same time, it opened the opportunity and challenge to think anew about the relation between ontology and epistemology. The main answer that emerged shortly afterwards, this was the other big progress in this field in the 70's, is two-dimensional semantics. It was certainly foreshadowed, but not clearly articulated in the papers of Kripke and Putnam. In a way, all philosophers of language at UCLA worked at that project at that time, culminating in Kaplan's (1977) essay. A different, but equally important interpretation was offered by Stalnaker (1978).

According to two-dimensional semantics, each referring phrase is evaluated, i.e., assigned an extension, along two dimensions, relative to possible contexts (of utterance) and relative to possible indices (of evaluation). Contexts are to be conceived as doxastic alternatives, this was Stalnaker's insight, and thus represent the epistemological dimension, whereas indices represent the ontological dimension. Accordingly, there are two kinds of intensions: horizontal intensions (= Chalmers' (1996) secondary and Jackson's (1998) C-intensions) as functions from indices to extensions (at a given context) and diagonal intensions (= Chalmers' primary and Jackson's A-intension) as functions from contexts to extensions which are defined as diagonalizations of the two-dimensional schemes (= Kaplan's characters) associated with the referring phrases. Horizontal intensions may well be called objective, ontological meanings, and diagonal intensions play the role of subjective, epistemological meanings, which thus diagonalize objective meanings. Hence, diagonalization is, I am convinced, the key to understanding the relation of ontology and epistemology.

This is the rough scheme. It is obvious, though, from my reference to the divergent origins in Kaplan's and Stalnaker's papers and to the still further diverging continuations in Chalmer's and Jackson's work that the proper understanding of two-dimensional semantics is not fixed or even codified, but rather contested and unclear. The situation is aggravated by the fact that there is a lot of philosophical terminology, rigidification and derigidification, response-dependent and response-independent concepts, etc., all of which allude to two-dimensional semantics in some form usually, and unfortunately, left implicit. Naturally, I for my part like best the systematic interpretation elaborated in Haas-Spohn (1995) and Haas-Spohn, Spohn (2001).

What I would wish, then, is a wider and profounder debate about the proper understanding of two-dimensional semantics and about its various philosophical applications, since, as I have indicated, it is not just semantics which is at issue, it is rather the deep relation between ontology and epistemology. Therefore, the issue is bound to have wide-ranging consequences.

For instance, the insight that the usage of the notion of an intension before 1970 and even afterwards has been systematically ambiguous is highly revealing. Does the meaning of a scientific term depend on the theory stated with its help? Yes, as empiricists from Carnap to Feyerabend have emphasized in different ways, provided meaning is understood as diagonal intension; no, as Putnam has fairly objected, provided meaning is understood as horizontal intension.

Possible worlds are ambiguous; according to Wittgenstein they are in some sense maximal states of affairs, and according to Lewis they are in some sense maximal things. Is there a tension? No, the two interpretations complement each other in two-dimensional semantics. Lewisian possible worlds are required as context worlds, and Wittgensteinian possible worlds are to be used as index worlds; it would be a mistake to assume an unequivocal use of possible worlds in two-dimensional semantics (a point in some way reflected also in Chalmers 2002).

Even the notion of truth is ambiguous. There is the correspondence notion of truth and its variants (about which there are most sophisticated debates). And there is an epistemological notion of truth described in variegated, though imprecise ways as coherentistic, as a limit of inquiry, as an evaluational notion, etc. Again, there is no conflict. Correspondence truth is truth at indices; epistemological truth, whatever its most adequate conception, is truth in contexts; and so both notions of truth find their place in two-dimensional semantics.

Might the ideal theory be wrong, as metaphysical realism claims? Yes, of course, in the sense of metaphysical possibility across indices; no, of course not, in the sense of epistemological possibility across contexts.

Or to end up my sample list of applications, take the notion of probability. We may well interpret de Finetti as having proved in his famous representation theorem that subjective probability is the diagonalization (= mixture) of possible objective probabilities.

Of course, each point would have to be argued most carefully. Still, my brief hints should indicate that two-dimensional semantics is capable of providing a comprehensive framework for dealing with the deepest problem in theoretical philosophy and for developing quite a number of most fascinating perspectives. Moreover, it appears to me that all the issues involved in stating the framework and its applications are sufficiently sharpened by the ongoing philosophical discussion, so that the attempt to definitely spell out the framework need not sink into confusion, but can immediately bear rich philosophical fruits. This is my hope at least, and this why I recommend this subject matter to utmost philosophical attention.

References

Blau, U. (forthcoming), *Die Logik der Unbestimmtheiten und Paradoxien*, Heidelberg: Synchron Wissenschaftsverlag der Autoren.

Chalmers, D.J. (1996), *The Conscious Mind*, Oxford: Oxford University Press.

Chalmers, D.J. (2002), "On Sense and Intension", in: J. Tomberlin (ed.), *Philosophical Perspectives 16: Language and Mind*, Oxford: Blackwell, pp. 135–182.

Dawid, A.P. (1979), "Conditional Independence in Statistical Theory", *Journal of the Royal Statistical Society, series B, 41*: 1–31.

Eells, E. (1982), *Rational Decision and Causality*, Cambridge: Cambridge University Press.

Haas-Spohn, U. (1995), *Versteckte Indexikalität und subjektive Bedeutung*, Berlin: Akademie-Verlag.

Haas-Spohn, U., W. Spohn (2001), "Concepts Are Beliefs About Essences", in: A. Newen, U. Nortmann, R. Stuhlmann-Laeisz (eds.), *Building on Frege. New Essays on Sense, Content, and Concept*, Stanford: CSLI Publications, pp. 287–316.

Jackson, F. (1998), *From Metaphysics to Ethics*, Oxford: Clarendon Press.

Kaplan, D. (1977), "Demonstratives. An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals", in: J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford University Press, 1989, pp. 481–563.

Kripke, S. (1972), "Naming and Necessity", in: *Semantics of Natural Language*, ed.by D. Davidson & G. Harman, Dordrecht: Reidel, pp. 253–355, 763-769; ext. ed. Oxford: Blackwell 1980.

Lewis, D.K. (1969), *Convention: A Philosophical Study*, Cambridge, Mass.:Harvard University Press.

Lewis, D. (1975), "Languages and Language", in: K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science, Vol. VII, Language, Mind and Knowledge*, Minneapolis: University of Minnesota Press, pp. 3–35.

McClennen, E.F. (1990), *Rationality and Dynamic Choice*, Cambridge: Cambridge University Press.

Nozick, R. (1969), "Newcomb's Problem and Two Principles of Choice", in: N Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel*, Dordrecht: Reidel, pp. 114–146.

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2000), *Causality. Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.

Putnam, H. (1975), "The Meaning of 'Meaning'", in: K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science, Vol. VII, Language, Mind and Knowledge*, Minneapolis: University of Minnesota Press, pp. 131–193.

Spirtes, P., Glymour, C., and Scheines, R. (1993), *Causation, Prediction, and Search*, Berlin: Springer.

Spohn, W. (1978), *Grundlagen der Entscheidungstheorie*, Kronberg/Ts: Scriptor.

Spohn, W. (1980), "Stochastic Independence, Causal Independence, and Shieldability", *Journal of Philosophical Logic* 9, 73–99.

Spohn, W. (2001), "Bayesian Nets Are All There Is To Causal Dependence", in: M.C. Galavotti, P. Suppes, D. Costantini (eds.), *Stochastic Dependence and Causality*, Stanford: CSLI Publications, pp. 157–172.

Spohn, W. (2003), "Dependency Equilibria and the Causal Structure of Decision and Game Situations", *Homo Oeconomicus* 20, 195–255.

Spohn, W. (forthcoming *a*), "The Core of Free Will", in: P.K. Machamer, G. Wolters (eds.), *Causation: Historical and Contemporary Perspectives. Proceedings of the Pittsburgh-Konstanz Colloquium VII*, Pittsburgh: Pittsburgh University Press.

Spohn, W. (forthcoming *b*), "A Survey of Ranking Theory", in: F. Huber, C. Schmidt-Petri (eds.), *Degrees of Belief. An Anthology*, Oxford: Oxford University Press.

Stalnaker, R.C. (1978), "Assertion", in: P. Cole (ed.), *Syntax and Semantics, Vol. 9: Pragmatics*. New York: Academic Press, pp. 315–32.

Stegmüller, W. (1952), *Hauptströmungen der Gegenwartsphilosophie, Band 1*, 7th edition, Stuttgart: Kröner, 1989.

20

Patrick Suppes

Lucie Stern Professor of Philosophy, Emeritus

Stanford University, CA, USA

Why were you initially drawn to formal methods?

Like a great many people interested in formal methods in philosophy, I was drawn already to certain kinds of formality in my early training in mathematics. I mean by "early training" that which I received in elementary and secondary school, especially in secondary school. By far the hardest math course I took as a high-school student was a course what was in those days called "solid geometry", meaning, of course, three-dimensional geometry. This subject is still a difficult one, especially when taught synthetically. Students of today who are progressing more rapidly through the curriculum will almost certainly, in an American high-school anyway, have a reasonably thorough course in calculus, but this was not the case when I was a high-school student in the 1930s. The second encounter with formal methods that I remember clearly was a course in calculus I took as a sophomore at the University of Chicago (1940–41). The instructor was someone who was himself then young, but became a very well-known mathematician—the topologist, Norman Steenrod. You would never have guessed from the way he conducted the course, and was not something that I learned until much later, that he was a brilliant mathematician. He seemed rather slow, extremely thorough, very thoughtful, but not very quick to give us answers. He would have long pauses when he wasn't sure of exactly how he wanted to say something. Thinking back upon it, he had a big influence because of the careful, detailed, and patient way he drew out of a group of reasonably bright but naive students the epsilon-delta methods for characterizing limits, instead of plunging right in with the compact notation used then in calculus courses. I was intrigued and involved by the methods he tried to teach us to give us a proper foundation for the calculus.