Davidson: Decision and Interpretation

Pol-Vincent Harnay* and Pétronille Rème**

* Independent Researcher
** Université Paris-Est, AME, SPLOTT, IFSTTAR, F-77447 Marne-la-Vallée

Abstract

Decision theory plays a central role in Davidson’s work. Based on the experiments led in Stanford during the 1950s, it is possible to track down the origins and the foundations of the unified theory of thought, meaning and action. The ‘wording effect’ and the omission of meanings undermine decision theory as a whole, hence the need to enlarge the basis of decision theory by integrating an interpretation theory that reflects mental holism more accurately.

Keywords: Decision Theory, Interpretation Theory, Mental Holism, Unified Theory.

Not much attention is usually paid to Davidson’s reflections on the role of decision theory in interpretation. These are, however, quintessential for anyone willing to understand Davidson’s holism and the way his theory of language and that of rational action intermingle.

Engel 1994: 111

1. Introduction

Very few studies have focused on Donald Davidson’s work on decision theory:1 the emphasis has rather been put on his groundbreaking work on action theory and philosophy of language.2 Nevertheless, Davidson is, in many respects, a major author of experimental economics, which developed in the United States in the 1950s. Besides, it is worth noticing the significance of decision theory in Davidson’s work—as evidenced by his numerous articles in which he deals with both his research at Stanford in the 1950s3 and his attempt, based on Richard Jeffrey’s research, to build a ‘unified’ theory of action and language that would overcome the weaknesses of decision theory. At least two approaches exist to gauge and put under perspective Davidson’s research work on decision theory.

1 Yet, Isaac Levi 1999 and Piers Rawling 2001 are worthy of quoting.
A first approach, through an emphasis on economics, would consist in examining into details the axiomatizations outlined by Davidson and his team in Stanford—including Patrick Suppes. The point would be to understand the theoretical foundations of their experiments as well as their role in the debates dealing with the theory of expected utility and, as a whole, their place in experimental economics. It is, however, another approach that will be tackled here, one that is conclusive with the interdisciplinarity of Davidson’s writings. The approach implies relying on the lessons and criticisms formed by Davidson himself after his testing in decision theory to better understand the embeddedness between decision theory and interpretation theory. I quote: “Theory of meaning as I see it and Bayesian decision theory, are made for each other”. The interest of such an approach lies, first, in its purpose that is to understand the origin of the unified theory of thought, meaning and action that was ardently defended by Davidson as of the 1980s. It was indeed from the experiments led in Stanford and the criticizing of decision theory that Davidson devised a more inclusive model of decision, one putting together decision theory and philosophy of language. Secondly, the approach sheds light on the consistency of Davidson’s work by offering a particular example of its ideas on mental holism.

In order to understand the interactions and imbrications between decision theory and interpretation theory, it is worth analyzing the central role of decision theory in Davidson’s thought through a focus (2. Davidson and decision theory: from early experiments to the “sophisticated” theory of reason explanations) on the experiments he carried out at the beginning of his career and above all by underlining the connections pointed out between decision theory and action theory. The different failures Davidson faced when he was an experimental psychologist are mostly due to skipping meanings in standard decision theory. Yet, the significance of meanings comes from the fact that choices are usually expressed verbally. As a result, when an experimenter, as often it was often the case in the 1950s, puts forward a behaviorist solution—according to which the mind is like a black box collecting data through stimuli, producing pieces of information in response—to the issue of measuring utilities and probabilities, he or she imposes a language. This language corresponds to the formulations of the options on which choices are made. A discrepancy can consequently appear between the meanings as perceived by the subjects’ experiments and the ones for an experimenter devising his or her protocol. The criticism is relevant since analyzing these meanings would offer an additional mental component to the experimenter, which would allow him or her to have a better understanding of the reasons behind decision. From that base, Davidson developed in the early 1980s a unified theory of action, language and interpretation (2. From decision theory to the unified theory of thought, meaning and action). Desires and beliefs are typically expressed verbally and Davidson reckons that the act of stating offers a piece of information that is directly linked to those desires and beliefs: one cannot understand what a person says if the former does not comprehend what the latter believes and desires. In other words, desires, beliefs, and meanings are connected by a strong mutual dependence that offers a significant image of the

---

4 For a detailed account of this approach, see Harnay 2008.
5 Davidson 1980: 158.
6 Davidson 1974a, 2001c: 236.
7 Davidson (1985) did use the term “base” to refer to his unified model.
mind. This image is both more relevant and significant than the standard behavior approach as its explanatory scope encompasses an additional datum, namely meanings: “The real problems of decision theory are problems of interpretation”.

2. Davidson and Decision Theory: From Early Experiments to the “Sophisticated” Theory of Reason Explanations

Decision theory was a particular component of Davidson’s career. It was the first project he worked on after his PhD on Plato’s Philebus. During the 1950s, Davidson published many articles about decision theory in which he detailed—along with Patrick Suppers—several experiments aiming at gauging the theory of expected utility’s empirical validity—this theory being the cornerstone of decision theory. Furthermore, his purpose was to offer an empirical interpretation of the theory, one that is testable.

Without explaining axiomatic into detail, it is relevant, though, to have a look on what Davidson learned from his experiments so as to understand why decision theory was so crucial for him. Interpreting the criticism expressed by Davidson on decision theory will be useful to fathom the emergence and the genesis of the problem of interpretation, which was his main topic of interest from the 1970s on and which fueled the afore-mentioned unified theory (2.1). In view of these analyses, it will be relevant to broadly introduce the connections between decision theory and action theory according to Davidson (2.2).

2.1 Lessons from Decision Theory

For the purpose of this article, I have decided to rely on the decision model that Davidson detailed in 1957 in Decision Making: An Experimental Approach. Such a choice is not trivial. After all, Davidson himself mentioned the model several times, underscoring with thoroughness its weaknesses and limits. In both cases, the point was either to draw an analogy with interpretation theory or to highlight mental holism. The experiments resulted for Davidson in various ways of tackling his unified theory: because of different effects—including an effect of formulation—undermining the experiments’ results (2.1.2), Davidson decided to promote the need to combine decision theory and communication theory. He added: “A radical theory of decision must include a theory of interpretation and cannot presuppose it”. Likewise, the behaviorist solution systemically showcased by the experiments in Stanford—which Davidson strongly criticized—is a way to pinpoint mental holism that was a recurrent topic of discussion among Davidson’s commentators (2.1.3). These two approaches will be analyzed after a brief overview of the 1957 model’s results (2.1.1).

---

8 Davidson 1999: 32.
2.1.1 The 1957 Model

Davidson’s input in 1957 took place as the foundations of the theory had already been laid, especially by the canonical model of Von Neumann and Morgenstern (Theory of Games and Economic Behavior [1947], corresponding to the second edition in which the appendix on the axiomatization of utility can be found). Within ten years, the epistemological and analytical stakes of their model were being discussed while the concepts used by Von Neumann and Morgenstern were being questioned—leading new axioms to be put forward, especially those by Friedman and Savage in 1948 and 1952.\(^{15}\) In this regard, the article by Friedman and Savage (1948) can be quoted as an attempt to empirically validate Von Neumann and Morgenstern’s expected utility theory. More precisely, it aims at collecting observations that deal with the behavioral choice of individuals facing risky outcomes in order to verify whether these observations are conclusive with what scholars call Von Neumann and Morgenstern’s “expected utility theory”. Finally, the goal is to examine the impact of the observations on the curve representing the utility function, especially in terms of aversion or appeal to risk.

Davidson’s theory, in line with that trend, is complex in many respects. The first respect has to do with its very purpose: as evidenced by the subtitle of his book written with Siegel and Suppes, Decision Making (1957), Davidson wanted to adopt “an experimental approach”. According to them, “no satisfying empirical interpretation of decision theory has been offered, therefore it is impossible to test it”.\(^{16}\)

The second one is linked to its method: Davidson, Suppes and Siegel expressed a whole set of theoretical hypotheses that were testable through rephrasing Von Neumann Morgenstern and Savage’s theory and that aimed at checking whether individuals maximized their expected utility by offering them several risky gambles. Based on empirical data, the scientists were able to devise a scale of utilities, evenly spaced, to verify the hypothesis of expected utility. What they did to conceive that scale was openly inspired by Frank Ramsey’s works and in particular his operational method\(^{17}\) allowing him to determine at the same time utilities and probabilities.

\(^{15}\) Even though the influence of Von Neumann and Morgenstern on Davidson and all was significant, the Stanford team clearly followed the path of Friedman and Savage’s works (1948), as proven by the attempt to obtain, thanks to experiences, utility curves similar to the ones hypothetically described by Friedman and Savage (for further details and a detailed account on the influences of the 1957 model, please see Harnay 2008).

\(^{16}\) Davidson 1957: 3.

\(^{17}\) Ramsey’s aim in Truth and Probability (1926) was to underscore the link between the subjective degree of belief that one has in a proposition \(p\) and its current probability. Furthermore, he focused on the way a degree of belief by an agent on a given proposition could be measured. If this agent applies a certain number of norms on rationality, then the degree of belief can be represented by a measure that is conclusive with the laws of mathematics of probability, according to Ramsey. Ramsey’s method consists of a ruler that represents numerically (thanks to different values) the way a person estimates the probability that something happens. The method consequently makes it possible to phrase a theory that would comprise several axioms requiring a rational behavior from the agent. From these axioms, one can calculate both the cardinal numbers of utilities (namely the subjective values of an individual on issues) and the extent to which an indi-
One major criticism of such a model expressed by Davidson is that some effects, like the formulation effect, undermine the results of Stanford team’s experiments.

2.1.2 The Formulation Effect

Before analyzing into details the content and the scope of the effect, two remarks have to be highlighted.

First, the harshest criticisms against the model in 1957 and broadly speaking against decision theory were only expressed by Davidson about twenty years after Stanford’s first experiments. All along his career in experimental psychology, Davidson commented upon the model merely a few times. For instance, Davidson and his colleagues discovered that winning or losing several times in a row made subjects respectively optimistic or pessimistic. As a result, this had an impact on the subjects’ subsequent responses to similar offers. The increasing (or decreasing) amount of money at play also influenced choices. This means that one must identify these distortions and find a way of avoiding contamination of all the choices made.\textsuperscript{18} Those are the kind of observations that are in Decision Making—observations that Davidson tried to address during the 1950s.

The second remark underscores the significance of psychologist Ward Edwards’ experiments (University of Michigan) in understanding the weaknesses of decision theory pointed out by Davidson. Edwards could be arguably viewed as the first experimental psychologist to describe in details all the variables that could affect subject’s choices, in one way or another. By showing how subjects systematically diverge from the objective model, Edwards opened the way to considering globally the psychological factors that influence choices. In 1954, Edwards talked about a particular effect: the “wording effect”. In Edwards’ mind, the wording effect corresponds to a test during which the experimenter changes the way options are verbalized by reversing the proposal that describes both gains and losses.\textsuperscript{19} Broadly speaking, the idea is to verify whether a reversal in preference results from the reversal in wording. In other words, does the wording influence the preferences expressed by subjects? From that point on, Davidson defended the idea that including a theory of interpretation in decision theory was necessary.

In Decision Making, Edwards is quoted twice, though for other reasons than the “wording effect”. In addition to distortions mentioned above, another issue expressed is the “recency effect”: even with the special die used, subjects got attached to specific nonsense syllables, for example “if the same syllable came up three times in succession”.\textsuperscript{20} To overcome the problem, the authors postponed

\textsuperscript{18} Davidson, Suppes, Siegel 1957: 53.

\textsuperscript{19} The options presented to the subjects were phrased as follows: “I toss a coin. If it comes up heads, I pay you $2.00. If it comes up tails, you pay me $1.00” (Edwards 1954), regarding the bets with a positive expected value. By presenting the losses before the gains while offering the same expected value, the goal was to know whether a change in the subjects’ choices occurred.

\textsuperscript{20} Davidson, Suppes, Siegel 1957: 54.
the payoffs and used three dice instead of one and “the die in use was changed after each toss”.\footnote{Ibid.}

In 1974, Davidson, in his article *Belief and the basis of meaning*, quoted Edwards through referring, directly this time, to the “wording effect”:

> There is not just an analogy between decision theory and interpretation theory, there is a connection. Seen from the side of decision theory, there is what Ward Edwards once dubbed the ‘presentation problem’ for empirical applications of decision theory. To learn the preferences of an agent, particularly among complex gambles, it is obviously necessary to describe the options in words. But how can the experimenter know what those words mean to the subject? (Davidson 1974: 147).

The argument is particularly relevant for the purpose of this article: Davidson established a parallel between decision theory and interpretation theory after criticizing the decision theory itself. He also mentioned the deadlock on meanings, whereas they are mental data necessary to reach a correct overview of the behaviors leading to deeds. This point will be developed in part 2.2.

The other main criticism of Davidson on decision theory deals with the behaviorist approach, which was systematically put into lights during Stanford’s experiments.

### 2.1.3 The Dismissal of Behaviorism

The solution advanced by Davidson, Suppes and Siegel to the problem of measuring utilities (desires) and probabilities (beliefs) is a behaviorist approach,\footnote{Davidson, Suppes, Siegel 1957: 12.} as explained by Davidson himself: “All we had to do was to give a clear behavioristic interpretation to ‘S prefers A to B’ and decision theory [...] became a powerful empirical theory, eminently testable, and palpably false”.\footnote{Davidson 1976, 2001c: 270.}

According to Davidson, behaviorism is problematic in its attempt to viewing mental conditions as simple physical conditions.\footnote{This has to do with the idea of anomalous monism: “The nomological irreducibility of the psychological means, if I am right, that the social sciences cannot be expected to develop in ways exactly parallel to the physical sciences, nor can we expect ever to be able to explain and predict human behavior with the kind of precision that is possible in principle for physical phenomena”, see Davidson 1974a.} A theory taking into account solely the physical aspect of mental conditions is indeed not worth considering, for Davidson: human behavior is part of nature (in this regard, “all mental events ultimately, perhaps through causal relations with other mental events, have causal intercourse with physical events”)\footnote{Davidson 1970, 2001c: 208.} but the idea according to which voluntary action can be applied to deterministic principles, like that of physics, is to be dismissed. One reason for the dismissal of strict psychophysics’ principles has to do with the holistic character of the cognitive field, which imposes to factor a growing number of components related to beliefs and motives for action. In other words, had this hypothetical physical theory of mental events rested upon such a foundation, it would have collapsed because of men-
tal holism that requires to add an increasing amount of pieces of information to be understood. The temptation to narrow decision theory to match physical sciences is thus doomed—impacting concurrently the possibility of empirically testing the theory, that is to say probing its expected results.

Nevertheless, though Davidson seems to condemn decision theory for the complexity—or even the impossibility—of testing the axioms of the field’s canonical models, he reaches a subtler conclusion:

I think I have an argument to show that the main empirical thrust of an explanation of an action in decision theory, or of a reason explanation, does not come from the axioms of decision theory, or ‘the assumption of rationality’, but rather from the attributions of desires, preferences, or beliefs (Davidson 1976, 2001c: 273).

Studying the links between decision theory and action theory will help going deeper into this idea.

2.2 Decision Theory and Action Theory

In the 1960-70s, Davidson imagined a theory of action that had many commonalities with decision theory, especially in its structure. Both theories shared the same goals by dealing with the reasons for acting, first, and the choices available for action, on the other hand: action theory aims at explaining the motives for an isolated deed while decision theory explains why an agent chooses one action among perfectly reasonable others:

Decision theory is a way of systematizing the relations among beliefs, desires, and actions. It does this by imposing a complex, but clearly defined, pattern on the way in which people’s beliefs and desires interact (Davidson 1997, 2001a: 126).

First of all, both action theory and decision theory rely on an analysis of the roles, respectively, of desires (preferences) and beliefs (probabilities). They are, then, based on the same principle of practical syllogism as proposed by Aristotle and consequently offer a “teleological” perspective.

Davidson’s decision theory, built on models and experiments from 1957-59, proved that desires and beliefs could be identified, singled out and measured—which cannot be done in action theory:

Given the idealized conditions postulated by the theory, Ramsey’s method makes it possible to identify the relevant beliefs and desires uniquely. Instead of talking of postulation, we might put the matter this way: to the extent that we can see the actions of an agent as falling into a consistent (rational) pattern of a certain sort, we can explain those actions in terms of a system of quantified beliefs and desires (Davidson 1975, 2001b: 160).

Action theory does not offer sophistication because of its structure and its underlying concepts, especially the one on rationality:

Two ideas are built into the concept of acting on a reason (and hence, the concept of behaviour generally): the idea of cause and the idea of rationality. A reason is a rational cause. One way rationality is built in is transparent: the cause must be a belief and a desire in the light of which the action is reasonable. But
the rationality also enters more subtly, since the way desire and belief work to cause the action must meet further, and unspecified, conditions. The advantage of this mode of explanation is clear: we can explain behaviour without having to know too much about how it was caused. And the cost is appropriate: we cannot turn this mode of explanation into something more like a science (Davidson 1974, 2001c: 233).

The principal difference between action theory and Davidson’s decision theory lies in the fact that the action theory he coined is a particular simple form of explanation by reasons that does not factor the way an agent makes a choice among several actions:

The discussion so far has been hampered, if not hamstrung, by my sticking to a particularly simple form of reason explanation, and this has prevented me from saying anything sensible about a number of problems, such as how an agent might be expected to choose among several competing actions, each of which is recommended by reasons he has. Similarly, no mention has been made of the effect of variations in the strength of desire, or degree of belief. The theory of decision making under uncertainty is designed to cope with these matters (Davidson 1976, 2001c: 268).

Likewise, action theory does not include the variations in desires’ intensity or the degrees of belief—as these issues have to do with decision theory.

For Davidson, however, decision theory is a refined action theory since it moves toward “scientific respectability”: “It gives up trying to explain actions one at time by appeal to something more basic, and instead postulates a pattern in behavior from which beliefs and attitudes can be inferred” 26.

Action and decision theories resort to a similar analytical reasoning. Action theory does not have a formal structure nor give the option to decide between two equally desirable actions. As for decision theory, it is static by dismissing meanings. 27. Davidson explains: “Decision theory purports to describe a static situation: the pattern of a person’s attitudes and beliefs at a moment”. 28 Earlier in the same article, he had shed light on the limits of a static theory on propositional behaviors: “how could we tell that subjects weren’t influenced in their preferences by the experiment itself—that their preferences weren’t changing as we went along?” 29

In order to overcome these shortfalls, Davidson introduced a new theory aiming at better assessing mental holism.

3. From Decision Theory to the Unified Theory of Thought, Meaning and Action

According to Davidson, decision theory, because of its formal nature and its normative mission, is silent about worldly matters; its abstract structure does not
provide meaningful interpretation about the terms it uses, such as “to prefer”.  

In other words, one of the main criticisms regarding decision theory in the 1950s deals with its avoidance of providing meanings. And it is precisely by adding these elements (2.1) that the unified theory emerged (2.2).

3.1 Adding Meanings

In his article A new basis for decision theory published in 1985 in a journal entitled Theory and Decision, Donald Davidson mentioned the recurrent difficulty of experimental decision theory: the experimenter acknowledges the analyses of meanings, or, more precisely, the way the subjects interpret the different objects of decision theory, such as gambles and their outcomes.

In his essay Expressing Evaluations (1984), which tackled directly such a difficulty, Davidson explained what the problems were:

Bayesian decision theories have a fatal drawback: they simply assume that an interpreter can tell what propositions an agent is evaluating or choosing between, or which interpreted sentences express the agent’s preferences [...] Decision theory begins with simple preferences between propositions; once these have been identified, the theory allows us to extract the beliefs and desires that went into, and explain, the preferences. But it says nothing about what determined the objects of the original simple preferences. Preferences are, of course, manifested in behavior in many ways. But this fact does not tell us how the content of preferences is fixed (Davidson 1984, 2004: 29).

This quote points out that not mentioning meanings in fact leads to at least two subsequent problems.

By leaving meanings out, decision theory overlooks an essential part of mental data that incite an agent to make a choice or a decision. By neglecting the analysis of meanings (or taking it for granted), decision theory is theory confined to a behaviorist approach whereas language is what ties the agent to people and to the surrounding world.

Another issue, correlated to the first one, can be summed up by using Davidson’s own words: neglecting the analysis of meanings implies “establishing the correctness of an attribution of belief or desire involves much the same problems as showing that we have understood the words of another”. More precisely, decision theory and interpretation theory account for tools of measuring interdependently mental contents, by resorting to similar methods. By ignoring interpretation theory, decision theory does without an additional tool that would provide, yet, a better understanding of mental contents.

This is why, according to Davidson, “what we must add to decision theory, or incorporate in it, is a theory of interpretation for the agent, a way of telling what he means by his words”.

Nevertheless, in order to do so does not mean adding to the 1957 model of decision ad hoc meanings but instead trying to interpret the meanings of proposals in actors’ eyes, while having access to their beliefs and desires.

30 Davidson 1999: 32.
31 The article is a revised version by Davidson 1980.
In other words, the aim is to conceive a theory, based on standard models of decision theory, integrating the analysis of meanings without, nonetheless, supposing in advance the existence of data to be explained: “this addition must be made in the absence of detailed information about beliefs, desires, or intentions.”

Analyzing meanings in decision is essential in view of the way experimenters usually assume that the words used by the subject and those used by the experimenter can be interpreted similarly. When choices are offered to the experiment’s subjects, language and its meanings are somehow imposed to the individual. According to Davidson, notwithstanding, imposing a language to the experimenter gets round the meanings that the subjects would give to the proposals and therefore neglects the information on the subjects’ behavior influenced by the analysis of these meanings:

To learn the preferences of an agent, particularly among complex gambles, it is obviously necessary to describe the options in words. But how can the experimenter know what those words mean to the subject? The problem is not merely theoretical: it is well known that two descriptions of what the experimenter takes to be the same option may elicit quite different responses from a subject (Davidson 1974b: 147).

For Davidson, however, the necessary interaction between decision theory and the interpretation of language goes farther. Both theories complement each other since they are tools to measure mental activities and deal with combined elements:

Theory of meaning as I see it, and Bayesian decision theory, are evidently made for each other. Decision theory must be freed from the assumption of an independently determined knowledge of meaning; theory of meaning calls for a theory of degree of belief in order to make serious use of relations of evidential support (Davidson 1980, 2004: 158).

As we have just seen, the criticism toward decision theory by Davidson echoes the various remarks made by Tversky in 1975 and displays many forms in Davidson’s several articles.

3.2 Toward a Unified Theory of Action, Language and Meaning

Integrating meanings allows Davidson to provide a broader theory where meanings, beliefs and desires are codetermined. This is in fact the unified theory of action, language and interpretation (2.2.2). In order to single out beliefs, desires and meanings, the experimenter turns into an interpreter (2.2.1).

3.2.1 When the Experimenter Becomes Interpreter

By placing the role of meanings at the very core of decision theory in its Jeffrey version, Davidson enlarges the experimenter’s initial part. In fact, its part will

---

34 Ibid.
35 Davidson 1974a, 2001c: 236-237. In a subtler way, one could say that meanings are supposed to be known as they come from the experimenter himself.
extend to that of an interpreter. As mentioned by Davidson, “theory of interpretation is the business jointly of the linguist, psychologist and philosopher”. In other words, by adding meanings and by considering them as seminal in individuals’ choices, the experimenter has to measure both the cardinal utilities and the subjective probabilities but also has to interpret the subjects’ utterances. According to Davidson, such a procedure would be as much a mental-measuring tool as that of Ramsey—though the former would be more precise than the latter.

Turning the experimenter into an interpreter requires a profound interaction with the subject. Particularly, the interpretation mentioned by Davidson implies a comparison of the desires, beliefs and meanings of the subjects with those of the experimenter. Doing so, Davidson offers a groundbreaking view in the debate about interpersonal comparisons, those that set the standards for everyone to compare each other.

Yet, the upgrade of experimenter’s role is only a detail shadowing the broader spectrum. As a whole, it is the entire empirical cornerstone—i.e. every mental information acquired—that is extended: meanings have to be processed at the same time as desires and beliefs. Decision theory has to integrate a theory of meaning.

3.2.2 The Unified Theory

More than twenty years after Decision Making (1957), Davidson offered a unified theory of decision and interpretation during the 1980s.

The original 1957 model described in the first part was enhanced and transformed through several articles written in the 1970s and 80s. In order to do so, this enhanced model was included into a broader “unified” theory by Davidson (1980). The term is justified by the very nature of theory’s purpose, a triplet (desire, belief, meaning) that encompasses a similar theoretical reasoning as well as a distinct approach since Davidson points out complex causalities between these concepts, which induce that one concept cannot be determined without the other two being determined simultaneously too.

Decision theory and interpretation theory aim at solving two problems: pulling apart the role of beliefs and desires for the former, separating that of beliefs and meanings for the latter.

As evoked earlier, Davidson discovered that decision theory assumes that one can identify and individualize the proposals that lean toward propositional behaviors like desires and beliefs. Nevertheless, the ability to identify the propositions underpinning an agent’s behavior is not to be separated from the one consisting in comprehending what the agent says. Davidson adds that one usually realizes what he or she wants, prefers or believes solely by interpreting those words. For the author, it is not easier to correctly establish the fact of attributing a desire to something than interpreting someone’s discourse. There is, still, a need to go farther and state that both problems are identical. As a whole, one cannot determine beliefs without mastering the language of an individual; and one cannot master the language of someone without knowing what he or she believes:

38 Davidson 1990: 318.
In order to interpret verbal behavior, we must be able to tell when a speaker holds a sentence he speaks to be true. But sentences are held to be true partly because of what he believed, and partly because of what the speaker means by his words. The problem of interpretation therefore is the problem of abstracting simultaneously the roles of belief and meaning from the pattern of sentences to which a speaker subscribes over time. The situation is like that in decision theory: just as we cannot infer beliefs from choices without also inferring desires, so we cannot decide what a man means by what he says without at the same time constructing a theory about what he believes (Davidson 1974, 2001c: 238).

The unified theory of decision and interpretation suggested by Davidson can be seen as a response to the experimental issues encountered by the Stanford team:

But stating theses mutual dependencies [between theory of meaning and decision theory] is not enough, for neither theory can be developed first as a basis for the other. There is no way to add on to the other in order to get started, each requires an element drawn from the other. What is wanted is a unified theory that yields degree of belief, utilities on interval scale, and interpretation of speech without assuming any of them (Davidson 1980: 158).

Among the expected results of such a theory, one could mention the enhancement of content due to the adding of meanings as raw data as well as the explanations available to clarify cases of potential irrationality.

Moreover, introducing meanings is a way for the theory to integrate a “dynamic” element: interpretation is built through a process of trials and errors. There are several steps before devising complete beliefs and meanings. In this respect, the principle of charity is the cornerstone that allows to maximize the agreement between the interpreter and the speaker.39

Guided by Richard Jeffrey’s work (1983), Davidson acted on his intuitions about the connection between the decision theory and the interpretation theory by offering a slightly different version of Jeffrey’s model:

We owe to Richard Jeffrey a version of Bayesian decision theory that makes no direct use of gambles, but treats the objects of preference, the objects to which subjective probabilities are assigned, and the objects to which relative values are assigned uniformly as propositions (Davidson 1980: 160).

Though Davidson decided to use Jeffrey’s model, it was not without making slight modifications. Indeed, conclusive with one of Davidson’s most important statements, one of the first proposals is that one should not take meaning for granted. This amounts to using non-interpreted sentences instead of proposals because from this point of view, the latter could be assimilated to meanings themselves.40

Decision theory is quintessential in Davidson’s work. The experiments he carried out in Stanford during the 1950s deeply influenced him on an epistemological and theoretical matter. The main problem pointed out by Davidson is that it does not include an interpretation theory of language. More specifically,

39 For a detailed presentation, see Engel 1994.
40 For a detailed presentation, see Harnay 2008.
the experimenter takes for granted the meanings bestowed upon the different options available by the subjects. According to Davidson, however, this is not obvious. Nothing indicates that the subject’s beliefs and meanings are analogous to that of the experimenter. As a result, the different criticisms about decision theory nourished a global reflection regarding the links between decision and interpretation leading to a unified theory of action and language. The underlying central point is that one cannot analyze nor comprehend decision without resorting to language as it is both a communication tool and an indicator of mental contents. By assuming such a connection, Davidson promotes a theoretical and methodological merging of disciplines that were heretofore divided, through an emphasis on the stakes behind such a merging.41

References

Davidson, D. 1974a, “Psychology as Philosophy”, in Davidson 2001c, 229-44.

41 We would like to thank Pascal Engel and Mélanie Harnay (Université Paris III Sorbonne) for their useful comments, their expertise and encouragement. A special thought goes to Audrey, Héloïse and Mathilde who gave us the strength to complete this adventure.


