

WHY THERE CAN'T BE A SCIENCE OF RATIONALITY D. DAVIDSON AND COGNITIVE SCIENCE*

Sofia Miguens
Universidade do Porto

Abstract

This paper aims at clarifying, from the perspective of Davidson's philosophy, the claim that there couldn't be a science of rationality and to assess its impact upon cognitive science, especially cognitive science research on rationality. First, I will identify a methodological assumption commonly made within that research. Then I will briefly describe the purpose of Davidson's Unified Theory of Thought, Meaning and Action and the role he attributes to rationality in it. Finally I will try to assess whether Davidson's view stands in the way of a science of rationality, or in any way questions the legitimacy of cognitive science research on rationality. I defend that although Davidson's view doesn't question such legitimacy it should keep us from assuming empirical research on rationality has the means to solve prior problems regarding the nature of thought, which bear on the way we conceive of rationality.

Keywords

Davidson, rationality, Unified Theory of Meaning and Action, cognitive science.

Introduction

Not every line of research within philosophy is of any interest to cognitive scientists. In fact, people working in cognitive science often have mixed feelings toward philosophy – they suspect something of importance might be going on, but the methods are obscure and the terminology strange. Common ground seems to be the nature of mind, but there's not much agreement on how one goes about building a theory of mind. Here I will take the work of a philosopher, D.

* This work results from an introductory seminar on D. Davidson and Cognitive Science which took place in the context of the ongoing Research Project *Rationality, Belief, Desire II – from cognitive science to philosophy* (POCTI/FIL/55555/2004, R&D Unit 502 – FCT) (Gabinete de Filosofia Moderna e Contemporânea, Instituto de Filosofia, FLUP).

Davidson - a philosopher whose work is not so close to cognitive science as for instance that of J. Fodor or D. Dennett -, and try to understand if it somehow bears on a particular domain of cognitive science research, namely research on rationality¹. I will specifically consider Donald Davidson's claim that there couldn't be a natural science of rationality, which may seem simply incomprehensible for those working in cognitive science. After all, empirical research on rationality (be it on reasoning, decision, the emotions, the evolutionary origins of modules of mind that account for biases in human reasoning and decision) is carried on as a matter of course in cognitive psychology², neuroscience³ and evolutionary psychology⁴. Research aims both at the personal level (reasoning and decision research is usually done using subjects' verbal performances and reports) and the subpersonal level, and so at what underlies subjects' reasoning and decision performance (as is the case with neuroscience's research on the emotions' modulating effect on means/ends management or evolutionary psychology's search for evolutionary reasons of those modules of mind which supposedly account for biases in reasoning and decision). In the literature one frequently come across such statements as that the subjects should make 'sound inferences' when they reason, consider only 'relevant evidence' for the fixation of beliefs, 'maximize expected utility' when they act, etc. That is, in fact, the reason why I am taking such research as empirical research *on rationality*. What I would like to know is whether Davidson thought such research – or the way conclusions are formulated - was in any way philosophically illegitimate.

Surely, a certain superiority and aloofness in regard to natural science's attempts at explaining human behaviour (especially those of psychologists) is not particularly rare among philosophers. One way to understand Davidson's claim would be to see it as a sign that he belongs to that lineage of philosophers, which includes for instance G. Ryle and L. Wittgenstein, who saw their own attempts at characterizing the mind as situated somehow above or beyond natural science attempts of explaining human behaviour. But is this really the case? In order to clarify, from the perspective of Davidson's philosophy, the claim that there couldn't be a science of rationality, I will do the following. First, I will identify a methodological assumption commonly made within empirical research on rationality. Then I will briefly describe the general purpose of Davidson's Unified Theory of Thought, Meaning and Action⁵ and the role he attributes to rationality in it. Finally I will try to assess whether Davidson's view stands in the way of a science of rationality, or in any way questions the legitimacy of cognitive science research on rationality.

¹ DAVIDSON, Donald, 2004, Could There Be a Science of Rationality? in Davidson 2004.

² Cf. for instance KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. (eds.), 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press. For overviews, see BARON, J, 1988, *Thinking and Deciding*, Cambridge, Cambridge University Press; EVANS, J. St. B. T., NEWSTEAD, S. & BYRNE, R. M. J. 1993, *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates.

³ Cf. for instance DAMÁSIO, António, 1994, *Descartes' Error – Emotion, Reason and the Human Brain*, New York, Putnam; DAMÁSIO, António, 1999, *The Feeling of What Happens – Body and Emotion in the making of consciousness*, New York, Harcourt Brace; DAMÁSIO, António, 2003, *Looking for Spinoza – Joy, Sorrow and the Feeling brain*, New York, Harcourt.

⁴ Cf. for instance COSMIDES, L. 1989, The logic of social exchange: has natural selection shaped how humans reason? – studies with Wason selection task, *Cognition*, 31, 187-276; COSMIDES, L. & TOOBY, J., 1992, Cognitive adaptations for social exchange in Barkow, J., Cosmides, L. & Tooby, J. 1992, *The Adapted Mind: Evolutionary Psychology and the generation of culture*, Oxford, Oxford University Press, 163-228; COSMIDES, L. & TOOBY, J., 1996, Are Humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty, *Cognition*, 58, 1-73; GIGERENZER, G., 1991, How to make cognitive illusions disappear: beyond 'heuristics and biases', *European Review of Social Psychology*, 2, 83-115; GIGERENZER, G. & HUG, K, 1992, Domain-specific reasoning: social contracts, cheating and perspective change, *Cognition* 43, 127-171.

⁵ DAVIDSON, Donald, 1980, A Unified Theory of Thought, Meaning and Action, in Davidson 2004.

1. Cognitive Science and rationality

What do cognitive scientists do when they investigate rationality? Generally speaking, they take *standards of correctness*, which they get from logic, probability theory and decision theory, as use them as models for empirical research. Once rightness criteria for thinking and acting are identified, data gathering follows, a great deal of it consisting in observing and describing subjects' performances in reasoning and decision tasks, and verbal reports on their own performance. It is assumed throughout that legitimate standards are available. Yet, taking abstract standards, whose formulations often involve mentalistic concepts such as 'belief', 'desire' and 'action', to somehow constitute *models* of agents' behaviour is assuming a lot. When for instance we take for granted that 'rational agents act, based on beliefs and desires, so as to maximize expected utility' we are assuming that (i) mentalistic concepts are somehow explanatory of behaviour and that (ii) the subjects understand language (since beliefs and desires are normally identified through linguistic utterances), and simply leaving it for philosophers to figure out what each of these assumption may imply. Surely, there are ways to argue for some of these assumptions – we may for instance consider that correct principles of reasoning and decision are *in our minds*, even if we don't always use them, minds being the level of cognitive systems described by beliefs and desires, and the reasoning and decision principles are an internally represented, chomskyan-style, system of rules, which would account for the subjects' competence. We would then have an analogous hypothesis, for rationality, to the one Chomsky introduced in cognitive science for grammar. From this point of view, any empirically described shift from standards of correctness could be ignored as irrelevant, a mere performance error, without the subjects being dismissed as 'irrational' (as from the point of view of Chomsky's linguistics one distinguishes a subject's grammatical competence and the subject's performance, and subjects' performances need not live up to their linguistic competence).

2. Davidson and a prior question

My point is that when we get involved in the type of research mentioned above, taking abstract standards of rationality as models for empirical research, questions are assumed to be settled which are in fact open. Among other things we are taking the status of mentalistic idioms - what we do when we think about ourselves and others by using 'beliefs', 'desires' and 'intentions' to think about 'actions', our own and others' – in the explanation of behavior to be a settled question. In fact, it should be noticed that not only do we use mentalistic idioms to think about ourselves and others but we cling to that form of explanation and are quite unwilling to drop the conviction that it is in those terms that we (our actions, our reasonings) should be taken and understood by ourselves and others. Davidson, who thinks there is something to such unwillingness, quotes⁶, in that spirit, a passage of G. Ryle's *The Concept of Mind* which is worth careful consideration, given the kind of cognitive science and philosophy of mind literature we read nowadays: «When we hear the promise of a new scientific explanation of what we say and do, we expect to hear of some counterparts to those impacts (like those of which physics treats), some forces or agencies of which we should never have dreamed and which we shall certainly never witness at their subterranean work. But when we are in a less impressionable frame of mind, we find something implausible in the promise of discoveries yet to be made of the hidden causes of our own actions and reactions»⁷. I think many of us, reading the Churchlands, or Damásio, in their more futuristic tone, about neuroscientific 'explanation' of action, do share this feeling, and believe, like Davidson, that there is something to the resistance to replacing mentalistic ways of explaining behaviour with hidden neurological causes of our actions and reactions. Where am I getting at? Empirical investigations

⁶ DAVIDSON 2004: 117.

⁷ RYLE, Gilbert, 1949, *The Concept of Mind*, 324-325.

about rationality often assume without further discussion that we already possess the rightness criteria for thinking and acting, i.e. that we have legitimate conceptions of rationality to start from. Now, if we may indeed avail ourselves to some definitions of what is rational, we cannot overlook the fact that we will only be able to apply such definitions to real cognitive agents, such as ourselves, via formulations involving mentalistic idioms. Thus, there's a question prior to evoking standards of correctness as models for real agents' cognitive performances, a question which concerns the status of mentalistic idioms, and, so, their explanatory role, their relevance in explaining behavior. This is a fundamental question for the philosophy of cognitive science and this is where I propose to bring Davidson in.

We should look on Davidson's positions about the nature of mind, action and meaning from a specific perspective: the situation Quine left philosophers of mind in, with his two ideas of naturalized epistemology and radical translation. Accepting both makes for a difficult situation: how can a naturalized approach to the mind be interpretive? Davidson's Unified Theory is an answer to this difficulty, which we inherited from Quine⁸. Like Quine, Davidson wants to start from overt behavior of humans, uninterpreted utterances, physical happenings, and reach the mental. Like Quine's radical translation, the Unified Theory attributes a role to rationality in such a project. It is in the light of this project that Davidson's claim that there couldn't be a science of rationality must be understood.

3. Davidson and Quine

As I said, I am assuming Davidson's Unified Theory faces the same problem Quine wanted to face with radical translation (to get to the mind – to belief and meaningfulness of utterances – starting from overt behaviour). We must bear in mind, though, that unlike Davidson and like behaviorist psychologists, Quine seems to have thought, at some time at least, that if there ever could be a serious science of human behaviour, it would be a non-mentalistic science, without any appeal to beliefs, desires and intentions. For Quine, mentalistic ways of explaining our own behaviour and that of others should be overcome by more effective explanations.

Because he believed that, Quine eventually found himself in the complicated position of having to account for linguistic behaviour in terms of dispositions. Quine's idea, a reduction project, was the following: we will reach the mental through language, taken as overt behavior. We must look upon language as a system of dispositions to verbal behavior, and we should look upon dispositions as physical states of systems. This is the way to reduction that the 'naturalized' approach to language and mind seems to force upon us. The problem is finding out whether what we end up with (language, conceived of as dispositions to behaviour) is reducible to the physical. Sometimes Quine expresses himself as if that were the case: dispositions are explainable in terms of physical characteristics of objects, in this case human organisms. Yet, in more recent writings, and given the extreme heterogeneity, compared with the 'disarmingly uniform mentalistic idioms', of both empirical content of ascriptions of belief and physiological mechanisms involved, Quine ends up recommending that we «take the best of what Davidson called anomalous monism», apparently dropping the reductionist intention⁹.

Now, unlike Quine, Davidson never thought there could be a 'serious', non-mentalistic science of human behaviour, going as far as producing a reductionist account of linguistic behaviour of humans. That's why their differences are so important when we're considering the need for mentalistic idioms in cognitive science. As is well known, Davidson's anti-reductionism about

⁸ FOGELIN, Robert, 2004, Aspects of Quine's Naturalized Epistemology, in Gibson, R. (ed.) *The Cambridge Companion to Quine*, Cambridge, Cambridge University Press; GIBSON, Roger F., 2004, Quine's Behaviorism cum Empiricism, in Gibson, R. (ed.) *The Cambridge Companion to Quine*, Cambridge, Cambridge University Press.

⁹ QUINE, W.O., 1985, States of Mind, *Journal of Philosophy* LXXXII: 7.

the mental is due to (i) normativity (of interpretation) (ii) causal character of mental concepts such as 'actions for which the agents have reasons' and (iii) externalism (the fact that mental properties supervene not only on the agent's physical properties but also on the world outside the agent). All those conditions of the mental make it impossible that there be laws and that human behavior be explainable on the basis of intrinsic traits of agents only (for instance the dispositions Quine speaks of). But if we cannot have a theory of bodies and their dispositions, does this mean the explanation of human behaviour is a lost cause? Not according to Davidson. Yet because he didn't share Quine's anti-mentalistic prejudices, Davidson conceived the project of going from the evidence we have, overt behaviour, namely linguistic utterances of humans, to the nature of thought, meaning and action in a totally different way.

4. Davidson, the Unified Theory and a science of rationality

Davidson called the most recent formulation of his programme the Unified Theory of Thought, Meaning and Action. He called it 'unified' because it brought together issues formerly dealt with separately in the theory of action and in the theory of meaning. Basically, the Unified Theory is a development of radical interpretation, itself a successor of Quinean radical translation, and still has as its aim to get from overt behaviour to mind and meaning. The Unified Theory's purpose is as ambitious as possible: understanding what makes humans intelligible to one another, taking only behavioural evidence, and so to understand what it is for words to mean something and for humans to think and to act. So, Davidson believes there is a way from the observable behaviour of humans to the nature of action, thought and meaning. Only his idea is, one goes along that way not reducing but relating («analysing from an equidistant point»¹⁰) the three basic concepts of belief, desire and meaning, taking them as coordinate elements for the understanding of behaviour.

In Davidson's own formulation, the theory seeks to determine three unknowns (belief, desire, meaning) from behavioural evidence, and not only two, like decision theory (which extracts beliefs and desires from overt preference behaviour of agents), and interpretation theory (which extracts meaning and belief from overt utterances of agents – statements manifest assent, as choice manifests preference). What brings Davidson to the Unified Theory? Essentially the fact that decision theory says nothing about meaning (even if agents preferences are identifiable only through linguistic utterances) and interpretation theory says nothing about preferences (and that becomes essential namely when the interpreter considers not only observation sentences but also theoretical sentences and the way the former are taken to support the latter). So Davidson thinks decision theory and interpretation theory are made for each other and sets out, with the Unified Theory, to bring decision theory into interpretation theory. Basically, he thinks that to get to meaning and belief from utterances, holding true is not enough as evidence: degrees of confidence in the truth of sentences and relative desirabilities attributed to the truth of sentences must also be taken into account. So if in radical interpretation the starting point was *holding true*, in the Unified Theory the starting point is *preferring-true* (one sentence to others), and it is to account for that that Davidson brings decision theory to bear upon interpretation theory.

What we need in order to understand the question we started from (could there be a science of rationality?) is to understand the role rationality plays here, i.e. the way rationality comes into the Unified Theory. We will have the answer we are looking for if we ask the following question: where does the Unified Theory get what it needs to solve for the three unknowns (belief, desire, meaning) taking only overt behaviour as evidence? And the answer is: from logic, a truth theory,

¹⁰ «It is too much to ask that these basic intensional notions be reduced to something else – something more behavioristic, neurological or physiological, for example. Nor can we analyse any of these basic three, belief, desire and meaning, in terms of one or two of the others (...) A basic account of any of these concepts must start beyond or beneath them, or at some point equidistant from them all». (DAVIDSON: 152).

and decision theory. In fact, that's why the Unified theory has many of the characteristics of a science (such as describing an abstract structure, with properties which may be proved and making prediction – of intentional action - possible). But it also means that the very possibility of the Unified Theory rests, in Davidson's words, on 'structures dictated by our concept of rationality'. The components of this attempt at getting to belief, desire and meaning from behavioural evidence are all, admittedly, normative, and without them there's no theory. We are finally getting to what interests me here: the whole theory (of thought, meaning and action) rests on the norms of rationality, they suggested the theory, and give it the structure it has¹¹. For Davidson, this is the kind of theory of mind, meaning and action that we can have. And that is what we must understand about the Unified Theory to understand the question 'could there be a science of rationality?'

Now, is this way of dealing with the questions of meaning, mind and thought, the Unified Theory, a theory which rests on structures dictated by our concept of rationality, *scientific*? What does Davidson think? To start with, he thinks that the Unified Theory, although it is intended as a theory of what it is to think, communicate, meaningfully use language, in beings such as ourselves, and also as a psychological theory, it is not meant to be natural science or in any way compete with natural science. More specifically, it does not intend (although Davidson sometimes refers to it as being a 'theory of knowledge we already possess') to be cognitive science, and that's one of the reasons why it irritates people such as Chomsky (who has commented that what Davidson says about language says nothing about how a child acquires a language, or about what is innate in the syntax of natural languages). The truth is that the whole davidsonian programme has even been accused, by people interested in the science of mind such as Fodor and Chomsky, of being simply *unscientific*.

But if Davidson is elaborating a theory of a 'knowledge we already possess', and he is not doing cognitive science, then what is he doing? I think one thing we may take him to be doing is clarifying what we do when we investigate the mental and put forward, namely in cognitive science, explanations of behaviour which involve mentalistic idioms. And he is saying that we do not start out empty handed, and so we cannot think it will be investigations such as the ones I referred in the beginning that will ever, by themselves, give us something like 'the nature of rationality'. So, is a view such as Davidson's in the way of a serious science of psychology? What are its implications? Is it the case that, if we are davidsonians, we cannot be seriously involved in cognitive science research? More specifically, what should we think about lines of research such as the ones I mentioned at the beginning? At least, as I said already, we won't think that they will give us the key to the nature of rationality. This is simply because before, or beyond, those investigations we cannot give up asking questions such as Davidson's (how are thought, language and action, possible for humans?) and the answers to that sort of questions will not result from empirical investigations. For Davidson, only a theory such as the Unified Theory can do the job. And this means that although there is nothing illegitimate in cognitive science research on rationality, those investigations do have a problem: while they cannot account for their own assumptions, they are often looked upon as giving us more than what they in fact can.

¹¹ «The entire theory is built on the norms of rationality; it is these norms that suggested the theory and give it the structure it has. But this much is built into the formal, axiomatizable parts of decision theory and truth theory, and they are as precise and clear as any formal theory of physics. However, norms or considerations of rationality also enter with the application of the theory to actual agents, at the stage where an interpreter assigns his own sentences to capture the contents of another's thoughts and utterances. The process necessarily involves deciding which pattern of assignments makes the other intelligible (not intelligent, of course!) and this is a matter of using one's own standards of rationality to calibrate the thoughts of the other. In some ways this is like fitting a curve to a set of points, which is done in the best of sciences. But there is an additional element in the psychological case: in physics there is a mind at work making as much sense as possible of a subject matter that is being treated as brainless: in the psychological case, there is a brain at each end. Norms are being employed as the standard of norms.» (DAVIDSON 2004: 130).

Conclusion

I started out with the intention of understanding, from the point of view of Davidson's philosophy, the claim that there could not be a science of rationality, and then assess its impact upon cognitive science. I have, thus, been assuming all along that Davidson gives a negative answer to the question 'Could there be a science of rationality?'. But in fact, things are not that simple. Davidson does not propose a single, negative, answer but a double one: he thinks there can be a theory of rationality – the Unified Theory is in fact such a theory, and it has certain characteristics of a science. On the other hand, it has no claims to being natural science. Does this translate into a verdict of illegitimacy hanging over cognitive science research on rationality? I don't think this is the case: what we have is simply a remark about the status of such research. If what we're aiming at is understanding the nature of thought, language and action, if that is what we mean by a theory of rationality, and only that will satisfy our desire for a 'theory of rationality', then what Davidson is saying is that it will be only by developing a theory such as the Unified Theory and not by means of cognitive science research of the kind referred to in the beginning of my talk, that we will get it. Research on reasoning, decision, emotions, etc, assumes we already have answers to the questions Davidson is trying to answer. But Davidson can help us formulate questions which are vital for our understanding of rationality but are not being raised in empirical research, nor should they be. We may find it problematic that Davidson ends up defending that the only theory of thought, meaning and action we can have will itself be based on the norms of rationality. This certainly makes it complicated to look upon those norms as being themselves part of natural science, subject of hypothesis to be tested, confirmed or infirmed. But this does not have to be seen as problematic – for Davidson it is just how it is: as he says, with a brain at each end (and that is the case both in our understanding of one another as mental and in cognitive science) norms will be employed as the standard of norms¹². We will get ourselves into a difficulty only if, as it frequently happens, we do not take into account the difference between empirical research about the mind and the kind of problems Davidson addresses. Then we will find ourselves in the situation Wittgenstein diagnosed some decades ago: «the existence of experimental methods makes us think we have the means to solve the problems which worry us; but the truth is problems and methods pass each other by»¹³. My intention here was simply to make us realize why this observation should be generalized to cognitive science research on rationality.

References

D. Davidson, *Problems of Rationality*, Oxford, Oxford University Press, 2004

¹² DAVIDSON 2004:130.

¹³ WITTGENSTEIN, Ludwig, 1953, *Philosophical Investigations*, Oxford, Blackwell (II, XIV).