Rationality, Belief, Desire: a research programme

Sofia Miguens[1]

> «Some animals think and reason; they consider, test and reject hypothesis; they act on reasons, sometimes after deliberating, imagining consequences and weighing probabilities, they have desires, hopes and hates, sometimes for good reasons. They also make errors in calculation, act against their own best judgment, or accept doctrines on inadequate evidence. Any one of these accomplishments, activities, actions, or errors is enough to show that such an animal is a rational animal, for to be a rational animal just is to have propositional attitudes, no matter how confused, contradictory, absurd, unjustified, or erroneous these attitudes may be. This, I propose is the answer. The question is: what animals are rational?» Donald Davidson, *Rational Animals*[2]

1. The nature of Project *Rationality, Belief, Desire II – from cognitive science to philosophy* and its motivations. Assigning rationality and irrationality.

The articles collected here result from the research project *Rationality, Belief, Desire II – from cognitive science to philosophy* (POCI/FIL/55555/2004) and should be regarded as explorations of the issues into which the Project branches[3]. At the centre of our interests lie the various aspects of rationality. We take rationality to be a trait of cognitive agents. Cognitive agents are representation-guided systems, characterized by some goal-structure, which behave in an adapted way in their environment[4]. We assume that in order to consider them as such

---

[1] Principal Investigator of *Rationality, Belief, Desire II – from cognitive science to philosophy* (POCI/FIL/55555/2004)

[2] Davidson 2001: 95.

[3] The introductory book *Racionalidade* (Miguens 2004) was used as a guiding plan for the development of the current project (cf. questions for a philosophical theory of rationality, pp. 19-45)

[4] This definition is a starting point, and we are very much aware of the fact that it glosses over cognitive science discussions about the status of representations and goals. Our sole justification for this starting point is the level of the processes we are mostly interested in here, which are higher cognitive processes. J. P. Maçorano's article in this volume, though, goes some way into exploring the problem of the nature of representation, as well as the issue of relations between information and what we, from a mentalistic point of view, call belief.

it is not strictly necessary to evoke conscious awareness from the start. This way of approaching agents is common in cognitive science. In philosophy, by contrast, approaches to specific issues of practical rationality (rationality of decisions, rationality in action) and theoretical rationality (rationality of reasoning, rationality in the process of fixing and revising beliefs), often take conscious awareness of agents for granted. Leaving consciousness aside for the moment, it is important to notice that (i) talk of *agents* is necessary for formulating any questions concerning rationality, and (ii) taking certain parts of the world to be agents means taking them to have beliefs and desires, and thus describing them by means of *mentalistic language*. From the viewpoint of philosophy of mind – although such questions naturally extend to epistemology and metaphysics – the nature of such states and such language is an open question. Thus the Project's philosophy of mind background, which is expressed in its title by the purpose of jointly dealing with the nature of *rationality* and with the nature of *beliefs and desires*.

We believe it is very important, in order to understand rationality, to try to understand phenomena of both practical and theoretical *irrationality*. In fact, the wish to weigh accusations of irrationality addressed to agents, pieces of reasoning, decisions, institutions, under various circumstances, was a fundamental motivation for the Project. What is even more interesting with these accusations is the fact that they are very often followed by an appeal to Reason, which is supposed to make it possible to overcome the former flawed situation (as for instance when, in an argument, people don't understand each other, also when one tries to ground the status of laws, in juridical and moral contexts, or when one considers the progress of societies, or the quality of political decisions). It is in fact all too frequent to proclaim rationality or irrationality without feeling the need to know what is involved therein. But the truth is, it is easier to say what a valid argument is then to say what a justified belief, an appropriate decision or a rational creature are. It seems that we all want to be rational, in the sense that no one wants to be regarded as irrational: we do not want to hold unjustified beliefs, we do not want to be bad at reasoning and deciding, we do not want to act counter to our own best judgment. But why is that so? What does it have to do with our way of being human and with the kind of minds we are? Because the fact is, although we apparently wish to be rational, we admit that very often that does not seem to be the case – humans are certainly capable of acting against

their own best judgment[5], of believing things they have no reason to believe, of not believing that which they have reason to believe, of believing in contradictions, and so on. Again, why is that so? Is it that when it happens that we believe what we should not believe, or do what we think we should not do, this happens just as a matter of ignorance, as when we are unaware of a particular rule which would apply at a certain point in solving a given problem, or do such facts tell us more about the way our minds work?

In order to face these problems we have to have a picture of the volitional aspect of our nature. In the Project, the following questions were intended to conceptually take us apart as agents for that purpose, and have been the object of constant attention: What are desires? What are intentions? What are emotions? What is it that ultimately motivates us into acting? Are we psychological egoists, always motivated by self-interest? Is it that only self-interested action can lead to any form of satisfaction or happiness? What is the origin of selfishness? Can it, or should it, be in any way overcome in moral and rational action? Is it possible to know what we want without knowing what we, ourselves, are? How is our will, or our willing, structured, and how does such a structure relate to our self-representations as agents? Where does motivation come from on those occasions where humans seem to have less regard for self-interest than for other people's, or for society's, needs (need for justice, for instance)? Is a moral action necessarily and ultimately non-selfish or is it the case that for there to be motivation there simply has to be selfishness? How do selfishness and the emotions relate? Are emotions simply irrational? Is it really the case that in agents such as ourselves reasons can cause actions? Beyond those definitions available and commonly used[6], what is, after all, rationality in action?

These are in fact questions our first rationality Project (*Rationality, Belief, Desire – motivation for action from the viewpoint of the theory of mind*, 2003-2005) already dealt with. Project RBD1 had

---

[5] This is, of course, objectionable, and not every member of our group believes it is even possible (cf. Mauro & Cadilha).

[6] Namely (i) the instrumental definition (a rational agent is capable of recruiting the means appropriate to achieve the ends she pursues – of course, for that, ends have to somehow already be there, in the agent, prior to any action) and (ii) the idea of acting so as to 'maximize expected utility' (again, agents preferences, considered in deliberation, should be somehow previously defined and stable). In both cases, agents have beliefs and desires which enable them to consider things in terms of means/ends, utility, probability, etc.

the question of motivation for action as its main focus. Within that project we dealt with issues such as instrumental conceptions of rationality[7], the belief-desire model for the explanation of action[8], the nature of reasons for acting, the relation between reason and passions in more or less rationalistic conceptions of the will and of rational action[9], the specific nature of mental states such as intentions in contrast with beliefs and desires[10], philosophical theories of emotions (focusing especially on the cognitivism / non-cognitivism debate)[11], psychological egoism as, ultimately, the core of rational choice theory, psychological-philosophical foundations of economics, inasmuch as these involve rational choices of agents[12], etc. The first Project had thus a strong component of theory of action and of moral philosophy[13] and that is still the case in project RBD2.

The other basic motivation for the rationality Projects, besides the general interest in irrationality phenomena and in accusations of irrationality also concerned a specific theoretical need for assigning rationality[14]. It was a motivation of a more technical nature: we were interested in the problems faced by interpretation theories in the philosophy of mind. In fact, theories of mind such as those developed

---

[7] Madeira 2003ª.

[8] Madeira 2003a.

[9] Miguens 2003.

[10] Madeira 2003b.

[11] Mendonça 2004.

[12] Cf. Carlos Mauro, PhD dissertation. Although this work came to extend itself to other questions, it started by considering the philosophical-psychological foundations of economics, especially the concept of psychological egoism as the core of rational choice theory. Psychological egoism is the idea according to which people are always motivated by personal interest. Ultimately, this means that the rational agent acts always in virtue of self-interest, aiming at the maximization of expected utility.

[13] This was a consequence of the development of the project. From the start, we intended to have a broad perspective of the extensive literature on theoretical and practical rationality. Another objective was to find answers to the following questions, in ways which would orient future research. The questions were: (i) what motivates an agent into acting? (ii) what does a philosophical theory of rationality look like? what kinds of issues does it deal with? Here we have taken as references the works of S. Stich, A. Goldman, R. Nozick and S. Blackburn. Once these initial steps were taken, the research interest of the members of the group have naturally become more specific: some members of MLAG are currently interested mostly in philosophy of action and moral philosophy, others in questions concerning the nature of logic, still others in philosophy of mind (in topics ranging from interpretation theories, to theories of mind, emotions, mental causation, identity theories, etc.)

[14] We should speak not only of rationality assignments but also about arguments in favour of the impossibility of irrationality.

by W. V. Quine, D. Davidson and D. Dennett, have as their starting point a rationality assignment[15]. Here, the origin of the rationality Projects goes all the way back to my work on D. Dennett's Intentional Systems Theory[16]. Intentional Systems Theory – I'm using it as a general label for the various theses D. Dennett's theory of mind includes – is a Quinean theory of interpretation, and is committed to assigning to an agent, by default, beliefs which are mostly true, and inferences which are mostly rational. In the absence of such an assignment, it is simply not possible to take certain parts of the world as minds. In the specific case of Intentional Systems Theory, assigning rationality is a condition for the theories of representation, of consciousness, of action, and of personhood. So the whole structure stands or falls depending on the legitimacy and coherence of such a starting point. In the above mentioned work, I tried to explore the conditions and the implications of such an assignment of rationality. In particular, I was then interested in understanding whether that was done aprioristically, as seems to be the case in another, more well known, interpretation theory of mind, that of Donald Davidson. The answer was negative. Before saying why, and since Davidson's position is of central interest to us, I will consider it first. In his article *Could There Be a Science of Rationality*[17], Davidson defends the following thesis about the status of his theory of thought, language and action: «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

---

[15] The origins of this idea can be found in the way W. V. Quine considers the charity principle within radical interpretation (cf. *Word and Object*, 1960).
[16] Globally assessing Intentional Systems Theory as a set of positions on philosophy of mind issues is the purpose of Miguens 2002.
[17] Davidson 2004a.

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.»[18].

Davidson's idea is thus that traits assigned in interpretation are determined by formal theories, and that happens prior to any actual interpretation of another being. We can see circularity here, we can also consider that Davidson presents an argument for the impossibility of irrationality of what is to count as a mind. Before we shun such circularity, maybe we should stop to consider that things could not look very different if rationality constraints are built into the structure of mind and language – it is not as if we could step back and look at that condition from the outside and then describe it. As John Searle puts it, «we may intelligibly debate theories of rationality, not rationality.»[19]

Anyway, what could be an alternative to this aprioristic view? An alternative would be, for instance, to consider that it is the design of agents that gives the rationality assignment grounding, and that such design is a result of evolution by natural selection. Rationality simply is to cognition what adaptation is to life; both are cases of function and adaptation. In fact this is the idea behind Dennett's rejection of any characterization of Intentional Systems Theory as instrumentalist. He doesn't see his interpretation theory as instrumentalist because he thinks that what the interpreter does is not to project rationality but rather to recognize existing patterns, resulting from real design of agents[20]. Explaining design is a task for the theory of evolution by natural selection, which, in case we are considering mind-design, means reporting theory of mind to sub-personal considerations about agents. This line of thought involves a second argument for the impossibility of irrationality of agents, besides the davidsonian aprioristic argument[21].

In general, if arguments for the impossibility of irrationality stand, we should conceive of irrationality as a phenomenon within rationality, a phenomenon for which one should find a place.[22] As Davidson puts it, «The sort of irrationality that makes conceptual trouble is not the failure

---

[18] Davidson 2004a: 130.
[19] Searle 2001: xiv.
[20] For an overview, cf. Miguens 2006b.
[21] Interestingly, the issue of impossibility of irrationality comes back under another guise in the most recent work of Project members. Cf. Mauro & Cadilha, present volume, defending the thesis of impossibility of irrationality in the moment of action.
[22] Cf. Amen, in present volume.

of someone else to believe or feel or do what we deem to be reasonable, but rather the failure, within a single person, of coherence or consistency in the pattern of beliefs, attitudes, emotions, intentions and actions»[23]. Yet, it is not easy to avoid the temptation to regard as irrationality in reasoning or in the decision-making of others what *we* ourselves take to be irrational – in fact a great deal of empirical literature on the subject confronts this problem. Ultimately, the question is whether rationality and irrationality – to use J. Cohen's formulation[24] – can be empirically demonstrated. In other words, does it make any sense to assume that in empirical studies of rationality we start with no presuppositions at all about the rationality or irrationality of agents, and then find out through experience whether specific agents are rational or irrational? Or there is something wrong with this way of looking at things?

Some empirical studies of reasoning and decision have been followed by the conclusion that irrationality has been demonstrated. This would be supported by finding out that actual reasoning and decision do not conform to ideal standards, such as those of logics, probability theory, or decision theory. People simply do not think and decide by following such principles. They tend rather to use heuristic principles which simplify situations and are in general effective but also lead to persistent biases (this was one of the main points of the classic book by A. Tversky, Slovic & D. Kahneman[25]).

More recently, some authors have tried to avoid such conclusions about persistent irrationality[26] – that is namely the case in evolutionary psychology studies in which the results of irrationality test-cases well known in the literature, such as Linda the bankteller and the Wason selection task, are reinterpreted. Persisting biases are seen as resulting from adaptive characteristics of cognitive devices, which because they are adaptive should not be considered irrational.

Are we to conclude then that no skill, mechanism or component of human minds can, inasmuch as it is looked upon from the point of view of evolution, be considered irrational? This would be too strong a conclusion[27]. I think there is a question prior to the interpretation of the

---

[23] Davidson 2004 a: 170.
[24] Cohen 1981.
[25] Kahneman, Slovic & Tversky 1982.
[26] Cosmides & Tooby 1996, Barkow, Cosmides & Tooby 1992. Cf. for a summary Miguens 2004, pp. 84-88.
[27] Trying to make clear what is at stake when standards of rationality are evoked as models for empirical research, Samuels, Stich & Tremoulet 2003 consider the possibility

results of empirical research, and this is in fact where Cohen's thesis makes sense. According to this thesis, irrationality cannot be empirically demonstrated because normative principles such as those of logic or probability theory must not be considered as natural science hypotheses, which may be tested, confirmed or rejected. They are rather the very framework for the approach. In fact, this thesis rejoins the positions in philosophy of mind I have mentioned above, positions according to which we cannot consider anything as irrational without presupposing rationality – if there is such a thing as irrationality is should be regarded as a phenomenon within rationality. The problem is, it is obviously not clear which rationality we are talking about here, since it cannot be identified with the usual formal standards. It is not clear either how rational an agent must be to be taken to be a mind – too demanding constraints, such as having a perfect ability for calculation or a consistent web of beliefs, seem impossible to sustain.

I have claimed, based on Dennett's work and on reasons related to theory of cognition, that the concept of rationality as it is used in interpretation theories cannot be a deep or precisely defined one[28]. It is rather an agent-level notion, behaviourally based, tied to instrumental means-ends relations. It applies to the agent as a whole and is not even based on real representations, taken to be natural kinds (this would be a Fodorian view of what makes for the rationality of real agents). Such a concept of rationality may be indispensable to the theory of mind, but it is certainly not possible to simply identify it with standards of rationality such as those logics, probability theory and decision theory provide us with. That is why rationality in this sense is not liable to any precise characterization: It is a pragmatic notion, which should, thus, not be considered a label for some kind of intrinsic cognitive value. With this view[29], I did not intend to refuse or in any way deflate the problems of theoretical and practical rationality (problems concerning

of a Chomskyan style competence. If that is the case, questions of cognitive architecture and modularity should be considered. In between pessimistic theses (we are irrational) and overoptimistic ones (supposedly irrational performances are justified by the evolutionary history of the species), Samuels, Stich & Tremoulet try, in *Rethinking Rationality*, a text which was quite often discussed in Project's meetings, to defend an intermediate way. Stich & Sripada try to show how from an evolutionist point of view, evoked by some to keep accusations of irrationality away, it is still possible to make sense of the persistency in agents of currently irrational devices. I am especially thankful to Tomás Carneiro for studying and discussing these questions, introducing them to the members of the group, as well as for the translations of the texts.

[28] Miguens 2002: 510
[29] Miguens 2002.

what we should believe and what we should do), but rather point out that we are not justified in evoking rationality as some kind of secure and well-known ground, the kind of thing we hope for when we give it the last word in questions of thought and action. The minimal conception of rationality which the theory of mind needs and which is good for dealing with any cognitive agent is thus not to be identified with much more sophisticated and specific notions such as belief consistency, deductive closure or perfect inferential capacity. The problems of practical and theoretical rationality remain untouched by such a thesis, which is of a different level. So, rationality as a concept in use in interpretation theory of mind is not perfect rationality. This thesis must be explored, but anyway, it does not seem possible to enumerate aprioristically a set of true beliefs and inference principles without which we would not call an agent 'rational'. In his pragmatist theory of rationality S. Stich formulates this by saying that *there is no way to formulate a priori constraints for every possible rational agent[30]*.

Interpretation theories are, within the field of philosophy of mind, and given the role rationality plays in them, those we have been most interested in. Naturally, the status of rationality assignments is only one of the aspects which might make one doubt whether it is possible to sustain such theories. Interpretation theories are anti-reductionist[31], and there is a general question whether anti-reductionism is ultimately coherent. In theories which take themselves to be physicalist (that is at least the case with Dennett's) the unexplained residue which is the interpreter is a big problem: what is the interpreter? Where does the rationality assigned by the interpreter come from? J. Fodor[32], who is himself an anti-reductionist, but one whose anti-reductionism has a totally different form, blames Dennett's approach for being 'transcendental' (Davidson is often accused of the same sin). For Fodor that is definitely no compliment; it is rather related with an absence of explanation.

For Fodor, admitting of real representations is the first step which makes talk of rationality of agents possible: representations are more fundamental than rationality and are in fact the ground for explaining

---

[30] Stich 1993.

[31] Davidson is very direct when identifying reasons for anti-reductionism: these are (i) normativity of interpretation, (ii) causal character of mental concepts such as action, (iii) externalism.

[32] Cf. Miguens 2005, J. Fodor e os problemas da filosofia da mente. Fodor, Davidson and Dennett were our first references where it came to the question How should one go about doing theory of mind?

rationality from a cognitive point of view. There where Fodorian intentional realism places real mental representations, making a subpersonal explanation of the rationality of agents possible, taking it to concern computations of representations, a transcendental theory of the mind places nothing. This is not an easy problem: under the guise of a discussion about the starting point for the theory of mind what is at stake here is ultimately how naturalism or physicalism on the one hand, and normativity and subjectivity on the other could possibly stand together[33].

Besides the general question regarding the place of mind in a physical world, the question of rationality also led us to a question about the specific type of minds which are human minds and the specific kind of doings which are human actions – the approach to the nature of mind and action one defends is inevitably reflected in something which we, as humans, should care a lot about: a conception of what it is to be human. In the work I have been referring to[34], I raised objections to the underlying intellectualism of Dennett's theories of consciousness and personhood. An ongoing discussion in the current Project concerns a similar intellectualism in Davidson: Davidson bluntly states in «Rational Animals»[35] that small babies, like snails, cannot justifiably be considered rational creatures – only creatures capable of having concepts of belief and truth are capable of objective thought and thus deserve such title[36].

2. Methodological issues: cognitive science and philosophy.

The two motivations for the Rationality Projects identified above, both related to the status of rationality assignments, led us into the fields of philosophy of action, moral philosophy and philosophy of mind. A few methodological considerations are now needed, in order to understand how the Project extended to other fields. Inasmuch as RBD2 Project was itself conceived as having a focus on methodology, they have theoretical relevance as well. While the first Project centred on the

---

[33] The fact that this is the strating point of McDowell's Mind and World is one of the reasons why we were led to this author.

[34] Miguens 2002, Capítulo 4.

[35] Davidson 2001.

[36] I believe there is something wrong with Dennett's and Davidson's intellectualism concerning these issues – part of what is wrong has to do with not considering perception in theory of mind. Within the project – cf. interviews with Charles Travis – that's what some of us have been working on.

question of motivation for action, the second Project was conceived as having a focus on methodological questions regarding the relations of philosophy with cognitive science – that is the justification for the sub-heading «from cognitive science to philosophy». The field of cognitive science is very diverse, and debates in the philosophy of cognitive science range from issues such as nativism, modularity and the nature of representations, which do relate to our current interests, to others which are not so directly related to them, for instance those concerning connectionism. We used our philosophy of mind framework to delimit our concerns and decided that our general purpose, in considering the relations of philosophy and cognitive science, should be to understand the status of theory of mind in a framework of naturalized epistemology. Along the way, we intended to try to make clear what one means by naturalized epistemology. A possible understanding of naturalized epistemology is, of course, that epistemology should simply drop all normative questions. That's why we thought it would be especially interesting, since our interests were focused on the normative phenomena of rationality and irrationality, to try and see where naturalized epistemology leads. Is it the case that the study of rationality should simply be handed over to cognitive science?

To deal with such an issue we also had to try to make clear what makes cognitive science and philosophy, especially philosophy of mind, approaches to the mind different from one another, if indeed they are. We were fully aware from the start of the fact that contemporary philosophers understand the field of philosophy of mind in very different ways. Actually, simply trying to identify and compare uses of the idea of naturalized epistemology in philosophy of mind led us into controversies which the very nature of philosophy of mind as a discipline provokes. We took as starting point and as guidance the philosophies of Dennett, Fodor and Davidson, but did not restrict ourselves to their works. What we tried to do was to take them as offering concrete answers (and different ones) to the question 'How does one go about doing theory of mind'. Anyway, at present maybe only Davidson is still common ground and common interest to the members of the group[37].

I haven't yet mentioned our main contention about the relations between philosophy and cognitive science concerning rationality. We assume there is a philosophical problem of rationality, beyond the

---

[37] Cf. Cadilha, in the present volume.

related cognitive problems (such as those concerning reasoning, decision-making, etc). We think that cognitive science research about questions such as reasoning, decision, emotions, theories of mind, is an essential contribution for a theory of rationality. Still, we think cognitive science is not sufficient to answer all questions we identified in the guidelines of the project. We think that a theory of rationality should include (i) a description or characterization of the factors at play on occasions when agents move from certain beliefs to others, add or eliminate beliefs from their corpus of beliefs, or opt for a course of action from several alternatives, based on a set of beliefs and desires; (ii) a set of hypotheses about the way we decide about rightness criteria when we talk of justifiedness or rationality of beliefs and actions; (iii) a set of hypotheses about the reasons why we want to know (if indeed we do) if our beliefs are true and our reasoning and actions rational. And those are things that won't be found in cognitive science alone.

Some clues to the difference between philosophy's and cognitive science's approaches would then be, for instance, that (i) standards of rationality such as those of logic and decision theory may provide us with models in the study of processes of reasoning and decision – but it is still necessary to say why they apply, if indeed they do, (ii) simply describing processes of reasoning and decision-making is not enough to understand the nature of the prescription involved, (iii) descriptions of application of rules and principles do not yet say anything about the connection between subjectivity and normativity.

One formulation of the philosophical problem of rationality would thus be the following. If we take consciously aware human agents as they think and act, in certain occasions, according to certain principles, the problem is: why should such principles be used in such circumstances? And why those principles exactly? In the project we dealt with this question in terms of criteria of rightness (a term we took from A. Goldman[38]), and we took it that what was needed here to answer such questions was a criterium of rightness. Criteria of rightness should make the reasons for which we take certain rules, norms or principles, to be standards of rationality explicit. Here we were led to alternatives as different as the supposedly apriori nature of logical knowledge, the public nature of language-games based normativity, or the survival value of psychological processes which maximize the number of true beliefs of agents (characterized, for instance, within a

---

[38] Goldman, 1986.

reliabilist theory of epistemic justification). But if there are several candidates to criteria of rightness how are we to decide for one? And what is it that we are doing when we are involved in such decisions? Goldman himself deals with these questions in the context of a conceptual analysis of justification – but even if we do not agree with that approach, what is important here is that we clearly see that research on rationality inevitably touches foundational questions about the nature of thought and language, and, ultimately, the relation of thought to the world.

Another formulation of the philosophical problem of rationality we took as reference is due to Robert Nozick[39]: there is a philosophical problem of rationality because there are certain agents – ourselves – who not only use principles to think and to act, but also should decide which principles they should use to think and to act[40]. This formulation forces us to ask questions such as: what is the nature of such principles? Where lies their power? What makes us follow them? And above all, what decision principle should be used to decide about which principles to use?

Questions regarding principles are often formulated having explicit normativity and conscious agents in mind. Yet, Nozick himself considers the existence of a kind of normativity prior to that state; inasmuch as there are descriptions of rationality which are good for any agent, consciously aware or not, descriptions related to *free floating rationales,* there are ways things are supposed to be. If we turn our eyes in that direction, if we decide that what we take to be rationality should be considered from this 'bottom-up perspective', in order to understand its place in the world, and not by evoking formal standards, we will end up looking upon explicit normativity as something which only later, in some cognitive agents, in some kinds of minds, came to exist. And once it was there, it involved a search for reasons independent of an immediate aim – a care for reasons, for the quality of reasoning and decision-making that, in Nozick's term (1993), '*now floats free*' and should be explained as such.

Going back to the general question of the difference between philosophy and cognitive science when it comes to dealing with rationality, besides having tried to formulate it in terms of the relation between standards of rationality and criteria of rightness, and in terms

---

[39] Nozick 1993.
[40] Cf. Nozick 1993 and Bizarro 2003.

of a decision principle for the use of principles, it also soon became clear that we could not go about trying to understand phenomena of rationality and of irrationality independently of a general theory of thought, language and action. We saw that as a specifically philosophical task, which involved considering first person perspective (whether we call it self, subject, will[41]) and understanding. I said before that we initially took as guidance the work of three authors, J. Fodor, D. Dennet, e D. Davidson. Working on Davidson became quite important for us here, as we tried to deal with questions such as (i) the nature of first person perspective, (ii) subjectivity as understanding, as well as the linguistic nature of such an understanding, and (iii) the connection of subjectivity and normativity[42].

One thing should be clear: the need for this focus on the nature of first person perspective can be defended independently of the idea according to which philosophical investigations should remain exclusively aprioristic. In fact, the kind of approach we favoured in the rationality projects implicitly states that we do not think philosophical investigations should stick to an aprioristic methodology. It is certainly the case that not all cognitive science research is philosophically interesting, nor should it be. Yet, one can not fail to notice that certain kinds of research, such as those of development psychology and evolutionary psychology, are, by their very nature, quite relevant for philosophers who have an interest in certain kinds of problems, namely problems concerning the nature of mind, language and action. In this Project, that is the case of research on decision, emotions, joint attention, etc. It is important to try to formulate what constitutes such relevance, from a philosophical point of view and I believe it has to do with the following[43]: those that come to be, in the kind of rational agent that we are, the mechanisms of mind and rationality, result from evolution, and thus from the historical contingency of a certain kind of

---

[41] J. Searle calls it in a very expressive way, in the context of his theory of practical rationality, 'the gap' (Searle 2001).

[42] Cf. Miguens 2005 and Miguens in the present volume.

[43] I must thank Charles Travis for the illuminating ideas which he has provoked in me through the interviews which have accompanied the development of the project. In these interviews it becomes clear that this has been done starting from the discussion of Wittgenstein and Frege, and frequently considering and criticizing Davidson, McDowell and Fodor. As a result, a new dimension of 'history of contemporary philosophy' of thought mind and language has grown, largely exceeding the authors we started with; this happened in fact as it became clear that the questions about rational agents that we were posing could not be dealt with while avoiding other more general questions about thought and world and the nature of experience.

interaction with a world. They are, as Charles Travis[44] likes to put it, *species specific*, that is, they are not characteristic of any thinker whatsoever but rather of a certain type of thinkers, the humans, as result of a specific process of evolution. Our 'mental ways' are thus, in that sense, parochial and their products should not be looked upon as the product of just any thinker but as the product of a specific type of thinker. This singular character has epistemological, metaphysical and ethical implications, and provides us with yet another clue for understanding the different strategies of approach of philosophy and cognitive science to mind and rationality: working it out as such is a philosophical task.

3. Sketching some answers.

Throughout the Project's activities we tried to develop answers to the questions for a philosophical theory of rationality that were mentioned above. I have already said something about that – now I will just recapitulate some points.

We took it that facing the philosophical problem of rationality means not only identifying standards of rationality but also going beyond them, looking for criteria of rightness, or for the justification of the principles, and working out reasons to choose between alternatives there. In doing that, we kept in mind that the questions 'Why is it that we want to have true beliefs?' and 'Why is it that we want to think and act on a rational way?' may lead to different answers when they are (i) posed about cognitive agents in general (this is where hypotheses about evolution and survival are plausible) and (ii) when they are formulated from within the agent. From within we cannot exactly say about agents that they want to have true beliefs; it is rather that they find themselves having been endowed with a certain cognitive design and acting in a certain way (instrumentally rational, maybe maximizing true beliefs, etc…). One should then repeat the question: is it really the case that we want to have true beliefs and to think and act rationally? From his pragmatist point of view S. Stich would say that as cognitive agents we do not care for the truth of our beliefs or the rationality of processes, in that those are not our aims, our ends[45], and that – our aims, our ends, what we want – is what we care about.

---

[44] Miguens 2005.
[45] In his contribution to this volume Tomás Magalhães Carneiro analyses the way this idea applies to a theory of emotions.

This is where questions about the volitional aspect of our nature enter – where do our aims come from? What is the nature of our desires? Are they brute determinations of the kind of beings we are? Is it in any way possible to take over and control that which we desire? What is it that makes us desire or want? Must we choose, when identifying that which moves us into action, between humean desires and kantian principles? How are we to conceive the role of emotions here? How is exactly that something comes to be of any value to us? We are capable, to a certain extent, of taking over and controlling that which we find ourselves desiring and the mere fact that it is possible to claim that there are reasons for actions, and that we are in possession of them, and that they cause actions, goes in that direction. It is doubtful though that it is ever the case that desires step away, in a kantian way, to make it possible for rationality in action to be 'pure reason'.

But is it really possible to know what we do in fact desire? In order to answer such question, we have to understand not only how is it that what we desire can be known (or recognized) by us, but also how it relates to what we think we are, to our self-representations as agents. This brings in the question of self-knowledge, and here we must consider different views – anyway the problem concerns the unity of the mind of an agent, and this bears on the way we think about how beliefs, desires, intentions and emotions stand together in a mind (what kind of unity is there (if indeed there is unity)?, what rules upon what? when I, for instance, want something that I do not want to want – is this wanting still me, or not?).[46]

Should we ever think that beliefs, desires and intentions are the sole intervening factors in action? Shouldn't we consider that emotions and other non cognitive processes, such as proto-emotions, have effect upon behaviour and mind, and have a role in the decision processes? At least some lines of work within the project were based on the conviction that cognitivist theories of emotions do not account for the importance of such non-cognitive processes, and so do not appropriately account for the way we decide and act[47].

The answers to the questions above must make sense within a general view of mind and thought, and there some decisions must be taken about how explicit a level we are dealing with. Is the attribution of rationality and irrationality totally within our power, or the very act

---

[46] The problem is partly dealt with in the article by Amen, in this volume.
[47] Carneiro, in the present volume.

of taking a mind as a mind – which we, being the kind of being we are – do anyway – has as its condition rationality assignment? According to interpretation theories, this is the case, and so irrationality is a phenomenon within rationality, and it simply does not make sense to try to empirically refute rationality.

Throughout the whole project, our basic idea was to think of rationality as a characteristic of real cognitive agents and their doings in the world, not taking for granted that we are already in possession of standards of rationality, but rather asking what, in such circumstances, do formal principles tell us about rationality processes. Our idea was that they provide us with models to think about such processes – and the need for models of phenomena is certainly common in science – but any extra step here, trying to claim more, would be risky. That is why discussions in the history and philosophy of logic could become very interesting – especially because they show us that logics may be conceived in different ways which correspond to different places and status for reasoning and thought.[48].

4. The articles.

I will now briefly summarize the contents of the articles included in this book, which result from ongoing research of the members of the Mind, Language and Action Group (MLAG) in the last year. As may be realized, by reading the articles, they stem from different positions regarding the problems identified above. One might even say that they express fundamental divergences among the members of the group, and indeed we take that to be a good thing.

In the first article, the contribution of Carlos Mauro e Susana Cadilha, *Why there cannot be an irrational action*, the authors defend that there cannot be contradiction in the moment of action, or the agent will not act at all. This is naturally, quite controversial, and much discussion followed the presentation of the talk, centering on notions such as revealed desire, revealed belief, differences between previous intention and intention in action, otiose beliefs, and the restriction of the application of the concept of rationality to an instant of action. Also, the authors general intention of deflating the importance of rationality in the production of moral norms was much debated and opposed by some

---

[48] Cf. Pinto, in this volume.

of the people present.

In his article, *L. Floridi and the philosophy of information,* J.P. Maçorano analyses the core concepts and main epistemological and metaphysical implications of the approach to the philosophy of information developed by Oxford philosopher Luciano Floridi. More specifically, he contrasts the concepts of information and data, pointing out the ontological differences between the two. Given such differences, Floridi's proposals are not compatible with a representation-based conception of knowledge. As J.P. Maçorano explains, Floridi defends an information-based concept of knowledge, taking information as data structured according to defined syntax and semantics. From a more general point of view, the relativist and pragmatist consequences of this position are explored.

In *Crença, triangulações e atenção conjunta,* Sofia Miguens continues previous work on Davidson's philosophy. As before it is assumed that only a general theory of the nature of thought, language and action may ground, and render systematic, the treatment of specific issues of the project. In Davidson's work it is possible to find such a theory. In this article, Davidson's views on triangulation, expressed in his last writings, are considered and assessed. These views change the former conception of radical interpretation, and bring intersubjectivity to bear on objective thought. The specific problem dealt with is the use of the concept of belief when considering non-linguistic agents. The article is also an attempt to explore the relations between philosophy and cognitive science concerning a specific problem, since Davidson's 'triangulation' is the object of empirical studies under the name of 'joint attention'.

Still in the context of the studies of the work of Davidson, Susana Cadilha, in *A teoria da acção de Donald Davidson e o problema da causação mental*, critically analyses some aspects of his philosophy, namely his theory of action and the ontology connected with it. Given the project's focus in agents and action, she focuses on mental causation, trying to make the implications of Davidson's position clear.

In his article, *Emoções e racionalidade derivada* Tomás Magalhães Carneiro considers the status of the non-cognitive background of agents in a philosophical theory of rationality, and also the possibility of finding a normative criterion of rationality for dealing with such issues. He discusses the implications of evolutionary psychology results – especially the work on emotions and on the rationality of emotion – for rationality theories. He considers the

specific proposals of S. Stich on the status of rationality, and also J. Searle's work on intentionality and the background. His main contention is that the rationality of proto-emotions is derived from higher forms of intentionality (involving conscious awareness).

In his paper *Davidson on Irrationality and Division*, Miguel Amen deals directly with Davidson's approach of irrationality. He defends Davidson from J. Heil's criticisms in Divided Minds. Davidson claims that in order to understand irrationality we should postulate a divided mind, while Heil poses objections to such a prima facie simple and direct claim. Even if a divided mind were sufficient to explain irrationality, it is not necessary (in fact is even superfluous). Miguel Amen defends Davidson from criticisms, while he also tries to correct Heil's interpretation of his theory.

Logics provides us with candidates to standards of rationality, namely for rationality in reasoning. Still it is not in any way legitimate to think of logics as simply revealing the rules any reasoning should follow (in fact, it is not even that simple to put forward a definition of reasoning, as something in contrast with transformations of information in a cognitive system). Before any claims are made about the relations between logics and reasoning, it would be helpful to make clear issues regarding the nature of logics itself. Work in the history and philosophy of logics is an important way in for that. That's what João Alberto Pinto's article, *Boole e Frege: matematização da lógica vs. logificação* is concerned with. He contrasts conceptions of logics as language and conceptions of logics as calculus, which can be articulated with different places for mind and reasoning .

I hope that reading these articles, which offer quite different perspectives and approaches to the philosophical problems of rationality, will prove to be illuminating and enriching.

Referências / References:

Ámen, Miguel, "Davidson on Irrationality and division", in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Barkow, J., Cosmides, L. & Tooby, J. 1992, *The Adapted Mind: Evolutionary*

Psychology and the generation of culture, Oxford, Oxford University Press.

Bizarro, Sara, 2003, "Robert Nozick e a natureza da racionalidade", *Intelectu* nº 9.

Cadilha, Susana., "A teoria da acção de Donald Davidson e o problema da causação mental" in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Caló, Susana, 2006, *A natureza histórica da cognição: debates filosóficos na teoria dos sistemas dinâmicos na ciência cognitiva*, Dissertação de Mestrado, Porto, Faculdade de Letras da Universidade do Porto.

Caló, Susana, 2004, "Do corpo e das crenças à acção: o mundo autista", *Trólei* nº 4

Carneiro, Tomás, 2005, "Para acabar de vez com o cognitivismo", *Intelectu* nº 11

Carneiro, Tomás, "Emoções e racionalidade derivada", in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Cohen, Jonathan, 1981, "Can irrationality be empirically demonstrated?" *Behavioral and Brain Sciences*, 4, 317-370.

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 statistitians after all? Rethinking some conclusions from the literature on judgment under uncertainty", *Cognition*, 58, 1 -73.

Davidson, Donald, 2001, "Rational Animals", in *Subjective, Intersubjective, Objective*, Oxford, Oxford University Press.

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

Davidson, Donald, 2004, "Could There Be a Science of Rationality", in *Problems of Rationality*, Oxford, Oxford University Press.

Galvão, Pedro, 2004, *"*Teoria da decisão, racionalidade e ética: o utilitarismo de Harsanyi",*Trólei* nº 4

Gigerenzer, G., Todd, P. & ABC Research Group, 2000, *Simple Heuristics that make us smart*, Oxford, Oxford University Press

Goldman, Alvin, 1986, *Epistemology and Cognition*, Cambridge MA, Harvard

University Press.

Kahneman, D., Slovic, P. & Tversky, A. (eds.), 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.

Maçorano, José Pedro, "L. Floridi e a filosofia da informação" in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Madeira, Pedro, 2003, "O que é a teoria instrumental da razão prática?", *Intelectu* nº 9.

Madeira, Pedro, 2003, "O que é o modelo crença-desejo?" *Intelectu* nº 9.

Madeira, Pedro, 2003, "A objecção de Nagel ao modelo crença desejo", *Intelectu* nº 9.

Madeira, Pedro, 2004, "Razão instrumental e razões para agir", *Trólei* nº 4

Mauro, Carlos, *A impossibilidade de irracionalidade na acção,* Dissertação de Doutoramento, em curso.

Mauro, Carlos, 2005, Somos ou não inevitavelmente racionais? *Intelectu* nº 11

Mauro, C. & Cadilha, S. "Por que não pode existir uma acção irracional" Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Mendonça, Dina, 2004, "Experience as reason: emotions and values in the construction of rationality", *Trólei* nº4.

Miguens, Sofia, 2002, *Uma Teoria Fisicalista do Conteúdo e da Consciência*, Porto, Campo das Letras.

Miguens, Sofia, 2004, *Racionalidade*, Porto, Campo das Letras.

Miguens, Sofia, 2005a, "O que tem a filosofia a dizer à psicologia?" Entrevista a Charles Travis, *Intelectu* nº 11.

Miguens, Sofia, 2005b, "J. Fodor e os problemas da filosofia da mente", *Intelectu* nº 11.

Miguens, Sofia, 2005c, "O problema do auto-conhecimento", *Intelectu* nº 11.

Miguens, Sofia, 2006a, "Why Can't There Be a Science of Rationality – D. Davidson and Cognitive Science", in Miguens, S., Pinto, J.A. & Mauro, C.E., 2006, *Análises / Analyses*, Porto, Faculdade de Letras da Universidade do Porto.

Miguens, Sofia, 2006b, "D. Dennett's brand of anti-representationalism", *Protosociology* 22.

Miguens, Sofia, 2006c, "Os problemas da filosofia da mente", *Diacrítica*.

Miguens, Sofia, no prelo, Charles Travis – entrevista nº 2 (Wittgenstein e Frege, pensamento e linguagem).

Miguens, Sofia, "Conceito de crença, triangulações e atenção conjunta", Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Morando, Clara, 2005, "Teorias da mentalidade: uma aproximação filosófica", *Intelectu* nº 11.

Nozick, Robert, 1993, *The Nature of Rationality*, Princeton NJ, Princeton University Press.

Pinto, J.A, "Boole e Frege: matematização da lógica *vs.* logificação"*,* in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Quine, W.V., 1960, *Word and Object*, Cambridge MA, MIT Press.

Samuels, R., Stich, S. e Tremoulet, P., 2003, "Repensando a racionalidade: de implicações pessimistas a módulos darwinianos", tradução portuguesa de Tomás Magalhães Carneiro, *Intelectu* nº 9.

Searle, John, 2001, *Rationality in Action*, Cambridge MA, MIT Press.

Stich, Stephen, 1990, *The Fragmentation of Reason*, Cambridge MA, MIT Press.

Stich, S. & Sripada, C., 2005, "Racionalidade, evolução e emoções", tradução portuguesa de Tomás Magalhães Carneiro, *Intelectu* nº 11.

Travis, Charles, 2006, *Thought's Footing*, Oxford, Oxford University Press.