Epilogue

Rationality in the Social Sciences: Bridging the Gap

Jesús Zamora-Bonilla

THE GREAT DIVIDE

Traditionally, there has been a 'great divide' in the social sciences between those theories based on a 'rational choice' approach to human behaviour, and those based on some kind of 'hermeneutic' approach.[1] The first side of the divide started in economic theory, as a formalisation of the idea of the 'economic man' of the Classical economists from the eighteenth and nineteenth centuries, though it has been 'colonising' other branches of the social sciences, particularly from the 1960s; one of its main features when seen at a distance, that is, when we don't enter too much into its details, is just its tendency to produce formal models: social outputs are computed as the result of individual choices, in such a way that both the individual choices and their possible interconnections are assumed to be subjected to some mathematical constraints (the 'assumptions' of the models). Hermeneutic approaches, on the other hand, don't like the 'scientism' that all this modelling transpires, and pretend to be (more) faithful to the qualitatively and essentially subjective essence of the human realities that constitute the social practices they study; these approaches tend to see their intellectual products more as a part of the 'humanities', than as a part of 'science' (in the Anglo-Saxon, straight sense of 'science' as either 'natural science', or research conducted according to the methods of natural science). Of course, there are other relevant differences, and, more importantly, there are other significant 'divides'; for example, there is the essential question of to what extent the social context or situation determines individual choices, and there are different answers to this last question in each part of the rational-choice vs. non-rational-choice debate. My aim in this final chapter is not, however, to map all these distinctions, but to concentrate on a very specific, but also very general and fundamental question, which is the conception of *rationality* that underlies the 'rational choice' and 'hermeneutic' approaches, for I think it is the one that has created the worst misunderstandings in the philosophy of social sciences, and it can serve better than other concepts to gather some ideas that have appeared in the preceding chapters. Still more specifically, my claim is that the main difference between both conceptions is connected to the attitude they have towards

the concept of *normativity*;[2] in a nutshell, rational choice theories have too few analytic resources to deal with the richness of the implications that this notion has in the social realm, but hermeneutic theories are too stuck to the prejudice that their interpretation of normativity capture it 'as it really is', that is, they ignore that 'normativity' (as 'rationality', 'interpretation', etc.), is just a 'theoretical term' in the social sciences, that can be logically analysed and refined; stated differently, hermeneutic approaches ignore that what they are producing is just *scientific models*. My goal is, then, to devise a strategy that can 'translate' the hermeneutic analysis of normative action into a 'formal model', in such a way that the relations between both paradigms are easier to see, and their virtues and drawbacks can more easily be compared.

NORMATIVITY AND REASONING

Rational choice theory (cf. Chapters 7 and 14) has often been criticised for purporting to treat human agents as 'mechanical', mere 'automata' programmed to maximise a function ('expected utility') on which those agents just have no say at all. I think this criticism is not fair, for rational choice theory is based on the assumption that it is not its own goal to *discover* what are the real 'utility function' (nor the 'probability function') behind real people's choices. The theory puts simply some limits to what an 'admissible' function is, a limit that reduces to logical consistency and the internal coherence of the preferences. What are the *real* expected utility functions of people is not a theoretical, but an *empirical* question, though our common-sense knowledge of how people are often allows us to introduce some additional assumptions (like diminishing marginal utility, a preference for more income instead of less, and the like).[3] So, rational choice approaches' eluding of the question of 'what do people really think and how do they really take their decisions' does not presume that human agents are automata, or something like that, but is the expression of the fact that, in general, those models simply take for granted that agents *have* 'rationally' thought and decided 'already', and analyse what *follows* from the assumption that those decisions are rational and mutually consistent.

Another usual criticism is that rational choice theory describes people as 'egoistic', only interested in the maximisation of their own welfare and as unresponsive to moral or ethical considerations. I also think this criticism is not fair, for, although it is true that in most rational choice models it is assumed that the agents' utility functions only depend on each individual's well being, this is not a *necessary* assumption of the theory, but a reflection of the idea that, in most 'economic' circumstances, people simply look out for their own welfare and that of their close relatives. But nothing precludes that, among the variables in the utility function's domain, one can include the 'moral sense of duty' the agent attaches to each possible action or each possible outcome.[4] The fact that economists and other rational choice social scientists do not usually include them is not a matter of principle (cf. Chapter 9), but it is probably because they consider that the hypothesis that 'people tend to do what they must do' has a little flavour of 'ad hocness'. Of course, a hypothesis being ad hoc does not entail that it is false, but scientists tend to prefer more parsimonious explanations (see Chapter 33); in this case, explaining people's complying with social and moral norms as an outcome of the interdependence of many individuals' self-interest (e.g. under the form of a network – Chapter 16) is clearly more parsimonious, because the self-interest assumption also explains many other facts about social life.

Nevertheless, there is indeed a grain of truth (or two) in these criticisms, but they are a little bit hidden. Regarding the first one, it is true that rational choice theory only attempts to identify the action or decision that is the *outcome* of the agent's reasoning process, and not the reasoning process itself (it simply

assumes it is 'rational', that is, that the process leads to the same outcome – or 'solution'– the rational choice model identifies, no matter how the agent manages to find it). But one can point to the fact that *reasoning is itself an action*, particularly (but not only) when it takes the form of *public deliberation* (see especialy Chapters 26 and 36). Hence, if we ask 'how does one reason *rationally*?', the theory has just *no* hint to give us; it seems to assume that people has some 'magical' capacity to discover the 'solution' of the equations that represent their choice problem, but it is obvious that, in general, people frame these problems in a way that is completely different from the representation of them that we find in microeconomic textbooks.[5] Frequent references to 'procedural rationality' (see Chapter 7) point to the fact that reasoning and deliberation is a real and costly activity, but these 'heterodox' approaches have concentrated in the investigation of 'non-fully-rational' thought processes (i.e. how do people think when they can not find out the 'rational' – that is optimal–solution), and my question is that rational choice theory has no resources even to describe the functioning of *rational* reasoning processes, because it does not even consider 'mental' processes as actions that are to be explained by the application of rational choice theory. After all, if the reasoning process leading to a 'rational' choice were an action whose *own* right outcomes (each of the *steps* of reasoning) came from the application of something like a rational choice model, this would lead us to an infinite regress, for we should also claim that the rational choice process leading us to the determination of the *steps* of the first reasoning is itself 'rational', and so on.

Regarding the problem of normativity, though it is true that rational choice theory is agnostic about the content of the individual's preferences, and this makes perfect room for the agent's having some *moral* preferences (amongst others), it is still true that the theory, which is based on the empiricist interpretation of preferences as mere *dispositions* to choose, or *revealed* preferences (i.e. $u(A) > u(B)$ if and only if the agent *would not choose B* if *A* were an available option), has no conceptual means at all to specify the difference between the *kinds of reasons* one may have to prefer an option to another one: it cannot tell whether you decide not to steal a jewel because a policeman is walking around (and you prefer to remain free instead of going to jail), or just because you think you *mustn't* do it. This might be regarded as a weak criticism, because, after all, rational choice theory does not also distinguish between *other* types of preferences (e.g. whether you prefer to study in order to get a job, or because you enjoy it, or in order not to upset your parents), and this is not taken as a problem, in general. But in this case there is a significant difference, because to say that you *must* do something is to say that you must do it *no matter what your preferences are*. A normative reason *being* a reason, hence, seems to consist in something different from being 'just' a preference. Actually, we can generalise this idea and apply it not only to *moral* reasons, but to all those cases in which we can say that one *has a reason* to perform an action, even if she chooses not to perform it, or vice versa. Hence, what rational choice theory fails to explain is what it is exactly *to be a reason*: the model assumes that the agent has a coherent repertoire of *dispositions* to choose, but it cannot answer the question about the nature of the *motivations* to choose, in particular, it fails to understand motivations *as* reasons, that is, it cannot answer the question about the connection between these motivations and the *reasoning* process we discussed in the previous paragraph.[6]

Normativity, intentionality, rule following, reasoning (at least in the sense of the interweaving of meanings that connects an agent's situation with her appropriate behaviour) and hence, meaning, are, instead, basic concepts in most of the non rational choice approaches in the philosophy of social sciences (cf., for example Chapters 6, 9 and 21–26). Many of these approaches have been based on the idea that the most fundamental

human capacity, and that on which society and culture are constructed, is *language*, an idea that has received a lot of conceptual support and analytical machinery from contemporary philosophy of language, mainly from the stream that flows from the 'second' Wittgenstein and the theory of 'speech acts' (see Chapter 5). I will take issue with this tradition in the following sections, by presenting a *model* of rational action which is based on the analysis of 'language games', and which, if successful, will allow us to show the mutual *consistency* of the hermeneutic and the rational choice options, by showing that the latter can be described as a *particular case* of the former. In particular, I want also to show that, though the hermeneutic approach is usually contrasted with rational choice theory on the basis of the latter's exemplification of a *nomological* kind of explanation of social facts, opposed to the idea of explanation as *understanding* proper of the former (see Chapters 1, 2 and 3), this is actually a mistake, for we can understand 'understanding' in a perfectly 'nomological' sense once the connections between meaning, normativity, and action are described in a 'scientific' model; a model, furthermore that allows us to connect in a 'natural' way the insights of hermeneutics with a *naturalistic* view of the human mind. So, one of the morals would be that hermeneutics is capable of absorbing the rational choice approach as a special case, to the 'cost' of recognising that there is only one basic form of scientific explanation, that is, nomological explanation, and one way of being 'scientific', that is, naturalism (see Chapter 4).

MEANING, ACTION AND DEONTIC SCOREKEEPING

In order to perform this task, I shall make use of one recent and powerful trend in the study of language and action. It consists in viewing linguistic actions, and particularly argumentation and persuasion dialogues, as governed by the submission to each speaker to certain *norms* that govern the connections between the *commitments* implicit or explicit in her past speech acts. There are several authors that have developed a normative approach of argumentation and language on similar lines (Stephen Toulmin, Chaim Perelman, Jürgen Habermas, Douglas Walton, Frans van Eemeren, Philip Pettit and others), but I shall refer to the exposition by Robert Brandom (1994), which I think is particularly suitable for philosophical discussion on the theory of rational action. My point of departure is a naïve psychological view of the ontology of mind and action presupposed in many social and philosophical theories: the mind contains *beliefs* and *desires*, which are somehow combined in more complex entities called *intentions*, and which in turn produce the agent's *actions*. Actions and other events induce changes in the world that may have the effect (amongst others) of bringing out new beliefs and desires. In this naïve view, the four referred types of entities are 'natural', in the sense that they are conceived as things (really, as 'natural kinds') that exist in human brains as an outcome of their natural functioning. Rational choice theory adds to this description three *theoretical assumptions*, intended both to capture the rationality of the agents, and also to give predictive and empirical content to the theory; these assumptions are, of course, that preferences are complete and transitive, that beliefs respect the axioms of probability, and that the performed action is the one that maximises expected utility (a mathematical combination of beliefs and desires). Although rational choice theory is often associated to an empiricist, nomological view of social science, the truth is that this view *is difficult to* 'naturalise', in part because of what we have seen in the previous section: how a 'natural' brain is able to implement the algorithms allowing it to systematically produce an outcome consistent with optimisation theory (while most often doing no calculations

at all!) is indeed a difficult question to answer. As we will see, a heuristic approach is, in fact, considerably *easier* to integrate within a naturalistic framework.

Robert Brandom, however, rejects this 'belief-desire-intention' furniture of the world (as well as its *prima facie* naturalism), and substitutes it for the conception that the elements and basic structure of human action are *fundamentally normative*, not 'natural'.[7] Brandom's description is, instead, something like the following: the basic units of our rational activity (which includes not only speech acts, but all kind of intentional behaviour)[8] are *commitments*, which can be of two types: doxastic (commitments to maintain the truth of a claim) or practical (commitments to perform, or not to perform, some particular action). These commitments (and entitlements)[9] can be subjected to two fundamental *normative attitudes*: they can be *undertaken* (by onself), and they can be *attributed* (to oneself or, more often, to others). The combination of all the commitments and entitlements undertaken by, or attributed to an agent at a given moment, is the agent's *normative status*. The most interesting aspect of Brandom's theory is that these statuses are essentially *inferential*, that is, they do not 'mean' anything in isolation, actually they are *nothing* in isolation (just a bunch of meaningless psychological humours or unintelligible noises, so to say), for their normative force is constituted by their being subjected to *inferential norms*. This means that undertaking a commitment consists in accepting *other* commitments, that is, those that 'follow' from the former commitment and the application of the relevant inferential norms (e.g. if I assert that my car has been stolen, I become committed to justifying my claim, to accept all the other claims that follow from the former, and to remove the inconsistencies with other commitments I have accepted before and that happen to be incompatible with the new one). Inferential norms are 'conceptual' norms in the sense that they are what gives content and meaning to the *concepts* employed in my claims: *meaning* something by some words consists in accepting what follows from the use of them. Of these inferential norms, some are 'intralinguistic', in the sense that they connect claims with claims (or doxastic commitments with doxastic commitments), but others connect language with other facts; basically, there are 'entry' rules, that confer a new normative status to an agent thanks to the happening of some physical event (paradigmatically, in the case of *perception*), and there are also 'exit' rules, that have practical commitments as their output (and, in the end, *actions*).[10] Lastly, interaction between rational agents basically takes place by what Brandom calls 'deontic scorekeeping', that is, each agent's taking into account the commitments and entitlements she is attributing to the others, and vice versa.

So, in a nutshell, as compared to the naïve psychology of the theory of action, now we have within the Brandomian framework doxastic commitments (and entitlements) instead of beliefs, and practical commitments (and entitlements) instead of desires, preferences or intentions. And, as compared to rational choice theory, *instead* of the assumption of logical coherence and mathematical optimisation, what we have now is the assumption that all the commitments of an agent *are inferentially articulated* according to the conceptual norms the agent is subjected to. As a representative of the hermeneutic approach, we see that the immense variety of inferential norms makes human action, as explained by Brandom's theory, much richer and more complex than what the 'toy' models of rational choice theory allow (an easy way of visualising this difference is to think of the fact that the actors habitating rational choice models hardly *talk* at all, they seem to live, so to say, in the times of silent movies);[11] but, of course, the cost of this richness is that a theory of action like Brandom's is nearly incapable of providing models that can be simple and manageable enough to be empirically testable through definite

predictions something that Brandom's theory is, on the other hand, surely not intending to do, of course.

SOME DIFFICULTIES FOR AN INFERENTIALIST MODEL OF ACTION

There are actually other things in Brandom's theory that impede its direct transformation into a 'fully fledged' theory of social action. The most important one is that there is no *explicit connection* between practical commitments (i.e., commitments to act) and actions themselves. One doesn't know, after reading what Brandom says about practical commitments, if the agent will *actually* do what she is committed to do, or will not; also one doesn't know what happens when the agent has incurred several mutually incompatible practical commitments (by the way, Brandom's theory also does not tell what happens when what are incompatible are doxastic commitments: the agent has there the *responsibility* to remove the inconsistency, but the theory does not say how to do it),[12] neither when the agent is simply 'entitled' to perform several incompatible actions (i.e. how does she *choose* in this case?; is it just by good-old-fashion utility maximisation?). Stated differently, in order to be employed as a template for social science models, we need to add to Brandom's theory some assumptions about the *causal efficacy* of commitments, or at least, we need some *empirical* criterion to determine what behaviours constitute the fulfilment of a commitment and which do not. In a sense, what we need is something analogous to what Paul Samuelson in the 1930s did about the concept of preference when replacing it by 'revealed preference', but for the case of 'commitment' or 'duty'; hence, we would need something like a concept of 'revealed duty'.

Another related problem comes from what is actually a philosophical virtue of Brandom's theory: in order to explain how 'objectivity' is achieved by a system based on the attribution of normative statuses, he claims[13] that we must look to the perspectival difference between undertaking and attributing a commitment: it is not that there exists a 'privileged' perspective (e.g. that of 'the' community) that serves to define the objective truth (which commitments are true and which ones are false), but that *every* 'subjective' perspective includes within its linguistic machinery the mechanisms for expressing the *conceptual* distinction between the commitments an agent really *has*, on the one hand, and, on the other hand, the commitments somebody *attributes* to that agent; analogously, these mechanisms allow the expression of the conceptual distinction between the fact that an agent is *committed* to some claim, on the one hand, and, on the other hand, the *correctness* of that claim. Basically, this means that, in order to attribute a doxastic commitment to somebody, you have to be able to distinguish the fact that she is committed to the claim, from the fact that the claim is true or not; hence, although you necessarily accept as true those doxastic commitments that you believe, you can also apply that perspectival mechanism to *yourself*, so recognising the *conceptual possibility* of your being *mistaken* in accepting that claim. This perspectival approach to commitments is, however, difficult to insert into an empirical model of action, because it creates a gap between the commitments one agent *actually undertakes* and the commitments the same agent 'really' has (not to talk of the commitments other agents actually attribute to her); the question is, what of these set of commitments are the ones we (as social scientists) must take into account as *causally* leading to action? A natural answer is that only *undertaken* commitments have *causal* efficacy, but if this is so, then, what is the role of 'real' commitments in the model?

A different way of expressing this problem is the following: when somebody accepts ('undertakes') a set of propositions (*P*), then she is 'committed' to accept *anything* (*F*) that follows from those propositions by the

application of the inferential rules she abides by. The question is that some of the elements of *F* will be *actually* recognised by the agent as being so, but many others will not, and hence probably she will not *undertake* consciously the latter, nor even implicitly (i.e. she may, mistakenly, even believe and express the negation of some elements of *F*). Hence, if we want to model the *dymanics* of normative scores, the problem is that we face a choice between including in an agent's score *all* the consequences of the claims she undertakes, or only those that she consciously recognises. The first option has the problem that it is obscure how an unrecognised commitment can have a causal effect on behaviour, whereas the second option ignores the fact that, after all, the agent *is* committed to those claims even if she does not notice it in any way, and we are left without any hint about how is it possible that the actual concatenations of arguments and reasons by an agent can be said to be *subjected* to inferential rules.

As a matter of fact, the two problems I have mentioned so far (what are the connection between commitments and actual behaviour, and what are the different explanatory roles of 'undertaken' vs. 'attributed' or 'real' commitments) are part of a single difficulty, that of understanding the connection between the 'normative' aspects of commitments and the 'descriptive' aspects of behaviour (see Chapter 22). But there is also a third important problem, which refers to *the nature of inferential rules*. From my point of view, the most severe shortcoming in Brandom's theory is that it gives no hint about where the conceptual rules come from. This is true both from the ontogenetic point of view (i.e. how does an individual *learn* what are the conceptual rules that regulate her use of words and the connection of words to perception and action, and in such a way that what he learns is not only an empirical correlation, but a 'necessary' – in the normative sense – semantic relation), and from the phylogenetic point of view (i.e. how have the actual rules that govern the 'right' use of words come to be exactly the ones they are). Furthermore, it is not clear whether Brandom equates all inferential norms with conceptual norms (i.e. rules that determine the *appropriate* use of concepts), or if there can be some inferential norms that are not 'conceptual' (e.g. *moral* norms, as 'you must not lie', can be interpreted as norms about practical inferences, as 'if you are asked something, then you must tell the truth').

HOMO DELIBERATOR: TOWARDS AN INFERENTIALIST MODEL OF ACTION

I think that all the difficulties pointed out in the previous section have to do with Brandom's insistence in keeping his theory as an analysis of a merely *normative* reality, and from his assumptions that the normative cannot be naturalised, and that norms are not reducible to regularities.[14] So, in order to transform inferentialism into the template of an empirical model for the social sciences, the following moves are suggested:[15]

1 The first thing to do is to become agnostic about a realm of norms that cannot be 'objects in the causal order',[16] and keep our attention *only* on the normative *attitudes* that transform a 'mere' psychological event into an attribution or an undertaking of deontic scores. That is, the idea is not to substitute 'purely' normative statuses for 'merely' psychological states, but to recognise *that some psychological states can consist in the recognition of a normative status*. In a more sophisticated way, we can also include *degrees* of commitment, for an agent can be more or less strongly committed to either a claim or an action. These degrees of commitment are, of course, psychological properties.
2 Given some actual commitments made by an agent (including the commitment to follow certain *norms*), we have to introduce some hypotheses about the *probabilities* of the actions the agent can perform. I will call these hypotheses 'behavioural laws'. The simplest assumption is that the agent *will* perform that action she is most strongly committed to, but other assumptions are possible (e.g. that the stronger a

commitment is, the higher is the probability of fulfilling it).[17] Actually, we can define 'rational' *action* as that which obeys some of these alternative assumptions; other types of behaviour, instead (i.e. when you do not tend to do what you are committed to do), would be 'irrational'. It remains an empirical question to what extent some people act 'rationally' or 'irrationally' in this sense. However, it is also true that, if we find that somebody seems to act against the commitments *we* attribute to him (or that we think she attributes to herself), we have to decide whether to revise our hypothesis about what behavioural law she is following, or our hypotheses about what are the commitments she is undertaking.

3 Commitments are inferentially articulated, which means that acknowledging some of them will lead you to undertaking others, according to the inferential norms you are accepting; since undertaking a commitment is an action, it will be taken according to some 'behavioural law'. Some inferential norms (those referred to as 'entry rules' above) command one to undertake certain commitments, given not (or not only) some previous commitments, but also certain circumstances as they are perceived by the agent. All this process of reaching some commitments from other ones, or from other things, is what *reasoning*, in a broad sense, consists of. Reasoning being 'rational' means, hence, that it follows a 'behavioural-law-for-the-undertaking-of-doxastic-commitments' which is 'rational' according to the sense explained in the previous point. This is coherent with recent 'dual-system theories' of cognition,[18] according to which, human reasoning works at two different levels: first, inferential jumps which are basically subconscious and 'intuitive'; second, 'reflective' inferential steps which are reducible to combinations of first-level jumps, but that consist in the conscious following of a rule. The capacity of being both *subjected* and capable of *mastering* the inferential links of your commitments, is the basic feature that makes of you, as a human being, an *homo deliberator*.

4 Psychological events, like having a belief or a desire, are also part of the 'circumstances' experienced by an agent. It is also possible, hence, that some 'entry rules' allow her to undertake a certain doxastic or practical commitment (with a higher or lower strength), under the assumption that she experiences certain psychological states that can be interpreted as beliefs or desires.

5 When the situation includes *more than one agent*, then the actions of each agent are part of the 'circumstances' of the others, and have to be taken into account according as how the inferential rules accepted by the agents command them to react to those circumstances.

In a nutshell, the basic elements of an inferentialist model of action are:

1 a description of the inferential (doxastic or practical) rules each agent accepts;
2 a description of the 'initial' commitments of the agents;
3 a description of their relevant circumstances;
4 behavioural laws indicating the probability that certain actions (including the undertaking of further commitments) are performed by the agents, given their previous commitments and their circumstances.[19]

The main problem with a model like this is, of course, that we have to introduce so many pieces of information that they can be adjusted ad hoc so to reach any conclusion we want (which is, obviously, a problem widely spread in the social sciences; cf. Chapter 34); so, the more constraints we are able to introduce into the elements *a* to *d*, the better for increasing the empirical content of the model. Rational choice models have an analogous problem (e.g. the utility and subjective probability functions can be manipulated ad hoc), though the number of independent variables in the case of inferential models can be much bigger (think, for example, of all the norms governing the meanings of each word, and all the practical rules regulating different circumstances). In some way, this difference is the core of the distinction between rational choice and hermeneutic theories: the former tend to *simplify* the social situations so that the modeller can draw some definite conclusions, whereas the latter tend to *replicate* the richness of the inferential articulation of reasoning processes. But my point is that this is only a difference of *degree*, not of kind: utility functions can in principle be made as complex and 'rich' as one wants (losing with that their predictive power), but inferential

models can also be made less and less complex (losing with that the 'sense of closeness' they give us).

To conclude this section, I will discuss briefly what can be the *logical* connections between inferentialist and rational choice models. The most important difference between them is that inferentialist models do not (or not necessarily) employ the concepts of utility and subjective probability, which were the result of *applying the hypothesis of logical coherence* to the more 'naturalist' concepts of desires and beliefs; instead, inferentialist models apply to the latter concepts the assumption that they are *inferentially articulated* by means of a network of practical and epistemic norms. The two models represent, hence, *two different theoretisations of the rationality assumption*, and it is hard to see whether both are mutually translatable. Nevertheless, once the psychological operations within an agent's mind or within a collective deliberation process have led the agents to some commitments, the link between the strength of the commitments and the actual choices could be described as a kind of maximisation (though I doubt this is necessarily so; see Note 17). Also, the process of reasoning can be modelled according to some Bayesian principles (though doxastic commitments are difficult to interpret systematically in terms of subjective probabilities). So, in principle, it is an open question whether we can replicate any inferentialist model by means of a rational choice model, or vice versa. If this were the case, however, I do not think it should be taken as a triumph of any of the approaches: it would only show that in both cases what we have are not the *real facts* themselves, but simply a couple of *abstract models* of the facts, and their mutual substitutability would entail that the 'great divide' is only an artifact of the history of thought.

More interestingly, and less speculatively, is the following: we can expect that there are cases in which agents *can* calculate (explicitly or implicitly) probabilities and utilities, and in these cases, it is possible that the *norm* of maximising expected utility is included within the inferential commitments of agents. Or there can also be cases where the actual inferential norms are mathematically equivalent (or very approximate) to some maximisation decision criterion. So, the inferentialist models allow that, in some cases, that *homini deliberatores* behave according to rational choice theory, or in a way that can be replicated by a rational choice model. This can be particularly the case when the inferential norms lead subjects to situations in which there is no action to which they are very strongly committed, but have, instead, several incompatible actions they are *entitled* to perform.

WHERE DO INFERENTIAL NORMS COME FROM?

In the previous section I have assumed that a social situation is defined, among other things, through the inferential rules accepted by the agents included in it. In order to build a scientific model of that situation, this is the appropriate strategy, of course, for we will be trying to analyse what actually happens in it. But this is open to a couple of questions. In the first place, it can be the case that 'we', as social scientists or philosophers studying that empirical situation, do not *share* exactly the same inferential norms (this can be obvious for practical norms, but it can also be the case for conceptual or epistemic ones). I don't think this can be taken as a problem; after all, it can happen that not all the agents *within* the same situation accept the same norms, and our model must take this possibility into account; so, why are the agents going to share exactly *our* norms? Of course, what is important is that the modeller is conscious of this possibility, and does not unjustifiedly project her own conceptual and practical norms upon her subjects.

In the second place, and more importantly, it is a scientifically relevant question *why* the rules admitted within a social situation are

the ones they are. It is unlikely that the agents have a systematic power to decide what will be the rules (in some cases –for example, the study of legislation – it can be the case, but this is rare), so they cannot 'change' them. But, on the other hand, it is also a fact that in different situations and places the rules are different, so rules *actually do change*. The naturalistic model sketched in the previous section gives us some hints to explain how this can happen. First, rules have to be learned, and are actually abstracted and conjectured from concrete circumstances; so, it is not necessary that in all circumstances the same rules arise, nor that all the agents learn exact 'replicas' of the same rules. Second, and perhaps more interestingly: our model recognises the existence of beliefs and desires *besides* that of doxastic and practical commitments; only the latter are inferentially articulated, that is, only from them we reach other commitments by applying or following the inferential norms we accept; but our, so to say, more 'basic' beliefs and desires, that arise in our minds as a result of merely *causal* processes (some internal, some linked to the external world),[20] are not necessarily *consistent* with the conclusions of our deliberations; actually, having certain rules (instead of others) will lead social groups to reach certain results (instead of others), for example, given the same 'external' circumstances, two communities with different conceptual and epistemic rules will end undertaking different doxastic commitments; it is, then, possible that one of these communities experiments with a higher degree of *cognitive dissonance*, in the sense that there are more and more severe inconsistencies between the doxastic commitments they have actually undertaken and the actual beliefs they have formed (because the latter do not arise *only* from those doxastic commitments, but also from other psychological mechanisms relating to perception, for example), and one community can also experiment with a higher degree of *dissatisfaction*, in the sense that the actions to which its members are committed don't actually lead to very *pleasurable* outcomes as often as in the case of the other community. My hypothesis is that those conceptual or practical norms leading to more severe cases of cognitive dissonance and dissatisfaction will tend to be replaced by others, perhaps in some 'Darwinian' way (see Chapter 20). Formal models for the evolution of norms within this framework (in which both deontic scores and psychological properties have a causal role) can be highly interesting to develop.

INNOCUOUS INDIVIDUALISM; HARMLESS COLLECTIVISM

The naturalist approach to inferentialism depicted in the previous sections allows us to understand in a very simple way one of the most enduring debates within the philosophy of social sciences: the individualism/collectivism question (see Chapter 8). In this debate, there is an undeniable bit of truth in the thesis that all social systems are constituted 'just' (or, 'in the end') by individuals; inferentialist models recognise this in accepting that 'original' attributions and undertakings of commitments are always psychological events, and hence, are the product of an individual's mind. This does not mean that there cannot be *collective* commitments: the inferential rules the agents abide by can perfectly allow the creation of this type of commitments, for the difference between an individual commitment and a collective one resides only in the different allocation of duties and entitlements characteristic of each case; as long as a particular commitment is necessarily 'linked' to more than one agent (in the sense that it can only be created or destroyed if it is created or destroyed for all the members of the group simultaneously), it will be a *collective* commitment, which can very well be different from the commitments of individual members.[21] Neither the social scientists nor the philosopher can legitimately establish a priori that inferential norms allowing the allocation of deontic scores in such a way are

not valid, that is, it is an empirical question whether those norms exist or not. It is also the case that the inferential rules can allow the creation of *collective agents* (like firms, clubs, armies, states and so on), which have *their own* deontic scores, not reducible to the duties and rights of particular individuals. Of course, the *undertakings* of entitlements and commitments by a collective agent can only take place *through* the actions of some individuals (for collectives have no 'original' psychological states), according to the relevant inferential norms, but this does not entail that those commitments and entitlements are *not* those of the collective agent (e.g. if a firm has the duty of paying a bill in a year's time, this is not equivalent to a list of duties of their employees or their owners, for all these people can be replaced by others in the next twelve months; the duty to pay is a duty *of the firm*).

Furthermore, inferentialist models also recognise, in a way that rational choice models do not, another usual claim of the critiques of individualism in the social sciences, namely, that the way in which the social situation is *normatively structured* is a determinant of the situation's outcome (cf. Chapter 22). According to rational choice models, only individuals' beliefs and preferences, together with the 'real' constraints faced by them (costs, resources and so on), determine the solution of the model; at most, some norms are recognised to have a causal role in the sense that some physically possible actions are prohibited by social or legal norms, though, in the 'deepest' applications of the rational choice model, these prohibitions are also explained as endogenous outcomes of the equilibria in the game between agents. In inferentialist models, however, rules are double-faced; on the one hand, they are constitutively normative, irreducible to combinations of individual strategies, and pre-existing to individuals (who have to learn them in many cases before they become fully capable of *rational* reasoning and action); on the other hand, they only work when they are *learned* by the individuals, who have to have a previous mastery of very complex cognitive abilities to be able even to interpret the behaviour of others as examples of rules. So, in this case as in the topic discussed in the previous paragraph, acknowledging that our models allow us to handle collective agents, collective commitments, transindividual norms and the like, does not force us to reject, nor to put within brackets, the *naturalist* stance on which the inferentialist models are built. I think the concept of 'institution' can represent in a more neutral way what inferentialism allows us to say about the natural realty of individuals and collectives. By an *institution* I mean a particular set of interrelated inferential norms connecting the normative scores of a group of agents (see Chapter 19). Some authors have defended a *communitarian* view according to which epistemic practices only make sense within groups or communities.[22] The schema depicted in the last paragraph is coherent to some extent with this communitarian view, but it also recognises that humans can only master a rich and meaningful system of inferential rules by learning it from others, and this is only possible thanks to the individual's ability to interpret the *behaviour* she is observing in others according to some patterns of thinking she must already have 'in her brain' (although the process of learning can substitute those patterns for others). The behaviour of other agents is always an 'external event' for you, and you cannot *see* the inferential norms your neighbours are following: the only thing you can do about it is to *guess* which norms they are, but in order to make sense of the others' behaviour as governed by some norms, you must first be able to discern whether your predictions about others' behaviour are fulfilled or not (e.g. how do you learn that 'no' means *no*?), and this ability is basically the same one that allows you to find out *physical* regularities in your environment and become angry or surprised when your predictions fail.[23] Your guessing the rules followed by others proceeds by trial and error, not necessarily until the actual inferential norms you

master happen to be *identical* with those of your neighbours, but only until the moment when you have *no further reason to revise them*, and this can happen in a social state in which each individual has some *different* interpretation of the inferential norms from the one other members of the group may have. This view of inferential norms as more or less variable is consistent with the observation that communities do not constitute completely homogeneous clusters with immense differences with other communities, but more or less diffuse sets with gradual differences (cf. Chapter 11). Hence, institutions, societies or cultures are more an idealised description of clusters of interconnected social practices, than some monolithic collective entities which exist independently of individuals. Nevertheless, our schema does not only allow us to describe institutions as constituted by the interrelated norms of single individuals, but, as I said in the previous paragraph, it also permits us to describe some institutions as collective agents which have their *own* deontic scores and inferential norms. Institutions in the first sense are *collective practices*, whereas in the second sense they are *collective bodies*.

PLAYING GAMES WITH HERMENEUTIC AGENTS[24]

How would a couple (or more) of *homini deliberatores* behave when confronted with a situation that can be depicted in game-theoretic terms? In principle, it seems that the strategic rationality assumed in game theory and the hermeneutic rationality described in this paper are very different: the former is about choices determined by payoffs, whereas the latter is about doing what one has to do. But I shall try to show that, though there are differences, they are not too large, and the ones that exist, are worth being exploited. Much confusion on the philosophical (and technical) implications of game theory has derived from the fact that, due perhaps to an excessive desire of simplification, or just to intellectual sleaziness, the usual description of a game (with the players and their possible strategies, on the one hand, and, on the other hand, the 'payoffs' got by the players depending on what combination of strategies is chosen) hides *three* different things behind the apparently simple concept of 'payoff'. First, there is the *outcome* (or expected outcome) of each player's choosing one particular strategy; this outcome is the, so to say, 'physical' or 'natural' *consequence*, or consequences, of the players *actions*. Usually, simplified games just state how much money each player gets in that case; in general, it is a description of how the gains and losses from the game are *distributed*. Second, there is the *valuation* each player makes of those outcomes or distributions, or the *wellbeing* she would get in each case, or, in general, the *psychological* description of the situation.[25] And third, there is the (usually counterfactual) *description of the choices* each agent would make on the basis of those valuations, that is, her *behavioural* dispositions (in rational choice theory, this corresponds to the maximisation assumption: the option with the greatest payoff *will* be chosen). The first two elements are usually also mixed under the single concept of 'utility function' (they correspond to the elements (3) and (5) in the definition of a 'strategic-form game', in Chapter 15): in fact, the concept of utility function, as it is technically employed in rational choice theory and much of social science, exclusively represents 'revealed preferences'; that is, the fact that $u(A) > u(B)$ means only that the player *would choose* A if she were offered a choice between *A* and *B* (if utilities are in *numbers*, these only reflect how this choice would depend on the numerical probabilities of getting the consequences of the actions *A* and *B*, if these consequences are uncertain). The utility function does not reflect per se any kind of psychological degree of satisfaction which might be greater if the agent chooses *A* instead of *B*.

The prisoner's dilemma		*Player B*	
		Left	*Right*
Player A	Up	3 3	4 1
	Down	1 4	2 2

Figure 1 The prisoner's dilemma

For example, when we examine the description of a game, like the prisoner's dilemma in Figure 1, and see that in one cell an agent gets a payoff of 3, and in another cell she gets 4, this *means* that, in describing the game in such a way, we are *assuming* that the agent would make a choice leading to the second cell if she had to opt between both. The question is, is that assumption *true*? If numbers are *also* assumed to represent thousands of euros, for example, then our hypothesis about the choice can obviously be true or false, depending on the player's valuations: if player B *actually* prefers the *distribution* of money given in the UL cell to the distribution in the UR cell, then she would not choose 'right', but 'left', if player A chose 'up', and this entails that the numbers reflecting the choice disposition of player B are *not* the ones given in the figure. As we saw in the second section, it can be the case that the real preferences of an agent include a concern for what *others* get, and so, we must understand the utilities or payoffs given in a game form as choices that agents would make *all-things-considered*, that is, as choices that take into account the player's total valuation of the *distributions* of gains and loses associated to each cell. We should give, in fact, two different tables in Figure 1: one representing the distribution of money associated to each cell, and another one representing the agents' valuations (that do not necessarily take only into account what an agent gets for herself) and their dispositions to choose (just if these dispositions are automatically identical to what follows from their wellbeing levels, what might not be: in case it is not, we would need an additional table, one representing the chances of behaving in one way or another taking into account the valuations depicted in the second table). Figure 2 shows the valuations of the outcomes described in Figure 1, corresponding to agents having something like a Kantian moral sense (numbers represent here the valuation each agent makes of the distribution depicted in the same cell in Figure 1; the higher the number, the more valued a distribution is; different valuation would have emerged if, for example, each agent's utility function comprises as arguments, with different weights, the outcome got by each player).

If the player is a *homo deliberator*, whose reasoning and decision processes are subjected to inferential norms, then this just means that it is those reasoning processes what we must take into account for deciding what numbers to write in the second (or third) table describing the game. In the same way as there is no universal law giving, for each possible set of outcomes, a definite utility function, we cannot either offer a general hypothesis about what inferential norms (doxastic and practical) each person will obey. But in both cases (choice models based on utility functions, or on deontic scores) we can make reasonably simplified assumptions, based on limited generalisations about what we empirically know about real people. For example, common descriptions of games assume that people would simply prefer more money to less (what is true in many cases, and false in many other cases); and,

The prisoner's dilemma for 'Kantian' beings		*Player B*	
		Left	*Right*
Player A	Up	3 3	2 2
	Down	2 2	1 1

Figure 2 The prisoner's dilemma seen by players abiding by Kant's categorical imperative

in the same way, we can assume that people examine many of the situations in which they participate, by trying to find out 'what they must do'. Usually, the task of finding out such a 'solution' to the game is carried out in a collective way, that is, by collective deliberation or 'communicative action' (cf. Chapter 26): each party offers reasons, and discusses or accepts the reasons offered by the others, till the point when the players agree in what is the 'cell' of the game that *must* be chosen, or find no further reasons to move each other's claim about what must be done. The practical inference for each player is, then, 'choose the option leading to that cell'. Technically, when there is an agreement on which is the 'best' cell, this usually transforms a non-cooperative game into a cooperative one.[26]

Naturally, two complications arise here. The first is that there is no guarantee that the collective process of deliberation leads to such an agreement, *even* if players are led by merely 'normative' considerations (e.g. they may have conflicting ethical values); in this case, in a game of more than two players it can be the case that some of them form coalitions that behave cooperatively within themselves, but non-cooperatively with the other coalitions. The second complication is that, as game theorists know well, a player's 'word' (or public commitment) is not always trustworthy, or, as we saw in above, the probability that an agent does what she is committed to do is not necessarily equal to one; so, the players have to take into account that, for each other agent, there is some probability that she fulfils her commitments and the inverse probability of her behaving in a merely strategic way.[27] So, what the inferentialist approach to rationality would suggest regarding the application of game theory to social situations is that these situations must be modelled in such a way that, in the first place, the *inferential norms* that allow the players to determine the values of each outcome must be made as explicit as possible. And second, a certain probability should be given to the hypothesis that players will behave according to deliberational principles, and the inverse probability that they will behave in a merely strategic way;[28] this entails that the game must be decomposed not only into a description of the outcomes and a description of the valuations, but in the latter case we must also give *more than one* description of the valuations: one that follows from the collective application of inferential norms, and another which follows from the individual application of these norms, in order to take into account the cases in which no agreement is reached, or cooperative behaviour does not take place from the beginning. The analysis of the game must take into account all these possibilities, but this is not an insurmountable difficulty for game theory, nor something really uncommon in it.

INFERENTIALISM AS SOCIAL EPISTEMOLOGY

To end this proposal, I will mention that the inferentialist approach can also be useful to illuminate some problems in epistemology, and particularly in social epistemology. I shall just to point here to a couple of examples.[29] First, the model allows us to understand the concept of 'knowledge' in a way that is not affected by the problems of its traditional definition as 'justified true belief'. Similar to the strategy of Craig (1990) of asking how the need for a concept like 'knowing' might have evolved in prehistoric times (instead of asking what 'knowledge' is), we may ask what is the inferential role that calling something 'knowledge' plays within deontic scorekeeping games. My hint is that 'knowledge' is basically a term of praise: we say (as agents *within* a game, not as philosophers examining the epistemic game from the *outside*) that somebody knows something when we are doxastically committed to the truth of what she says she knows (under circumstances that make us accept she is truthful, of course), and hence, to say that someone

'knows' is just a way of undertaking a commitment, or acknowledging a systematic way of inferring a commitment *for us* from a commitment *of her*. In a parallel way, you say that you *know* something when you think that you are *entitled* to 'pass' to the audience your own doxastic commitment to that statement, for example, because the evidence on which you base your opinion could be publicly displayed if necessary (and, *mutatis mutandis*, you say that you *believe* something when you don't think you are entitled to transfer your commitment in such a way to others, either because your commitment is not very strong, or because you are not capable of presenting the appropriate evidence).[30] So, saying that a commitment is a piece of knowledge means simply something about the *quality of the warrants* (according to what the inferential norms say that warrants *are*) that the speakers are able to display between them in the game of transferring commitments from agent to agent.[31]

The former is an explanation of what counts for an agent or a community to *take* something *as* knowledge. Another obvious epistemological question is how to *assess* this knowledge and the epistemic practices allowing us to attain it. An epistemic institution (i.e. an collection of interrelated inferential norms leading mainly to doxastic commitments) will objectively tend (under given circumstances) to produce commitments with some properties rather than ones lacking those properties; we can call *implicit epistemic values* those goals that an epistemic institution seems to be promoting (see Chapter 35 for an examination of the normativity issue within social epistemology). Of course, the agents can also have themselves an *explicit* view of which ones are the goals they are promoting through the application of their inferential norms. From a more or less simplistic perspective, we could say that, for the individual or the group considered, those claims are just the ones they think they *must* make, and so they don't need any *further* assessment than the one which is implicit in the constitution of that stock of knowledge (i.e. the assessment implicit in *their* inferential norms). This view, however, would be simplistic because agents can reach a *defective* situation even according to their *own* standards (e.g. they can have misapplied the inferential norms, according to their own interpretation of them). A normative assessment of the set of claims attained by those agents could consist, then, in showing what other claims *would have been better* according to *their own* standards. But there is still a further step we can take, for we can make an assessment of the inferential norms themselves; after all, some norms can conflict with others, pulling into different directions, so to say, and we can try to identify in the set of criteria employed by the agents some 'metanorms' that help to solve this tension; or we can find out some argument showing that some goals the agents actually endorse would be promoted much more efficiently with a different set of inferential norms. Naturally, all this applies as well to the assessment of *our own* epistemic practices. So, epistemology, when it touches the basic philosophical question of what is it that a statement has to have in order to *deserve* to be taken as a doxastic commitment, reduces basically to the question of what are the 'best' epistemic practices, or inferential rules, *we* could have.

This idea is also helpful when applied to another old dispute in epistemology, that between rationalists and relativists. For it opens a new bridge between the position according to which all knowledge claims are biased by interests and prejudices and the position according to which there are objective and absolute standards of truth and certainty. This bridge consists in recognising both that knowledge claims always derive from *contingent* epistemic practices and rules, that might in principle be different for different groups or contexts *and* that the social scientists (and the philosophers as well) are themselves agents *committed* to certain epistemic norms. Hence, the fundamental role of deliberation in the pursuit of knowledge would consist in asking the other

agents what would they *take* as 'knowledge', what ways of establishing a claim would they take as *appropriate*, what do they think would be persuasive for *them*, and afterwards engage in an epistemic practice that accords those standards. So, the goal of knowledge production is by no means that of establishing truths according to 'absolute' criteria, but to do it according to the criteria that *real* contenders would accept. One might object that this is too base a goal for science and philosophy, but this objection would be contested by simply demanding the standards to which the objector would defer.

The survey of approaches, theories and methods included in this handbook reflects, I think, how social sciences and their philosophical understanding are moving towards something similar to this deliberative paradigm. People working in different 'schools' are progressively more willing to use ways of reasoning that (at least according to their own understanding) approach those of 'rival' traditions, not least because of the flexibility allowed by new mathematical technologies (cf. Chapters 31 and 37), a flexibility that has contributed to blocking the old identification of certain schools with the application – or presumed impossibility of application – of a given corpus of maths. If this is just an optimistic illusion produced by the recent wealth of new software, or something that is going to change radically the way that social sciences are practiced, is something that only time will tell.

NOTES

1 I am using the term 'hermeneutic' here in a rather wide sense; I admit that not every non-rational-choice approach in the social science is, literally speaking, a 'hermeneutic' one, but I think that the 'hermeneutics' label represents, particularly for philosophers, a quick way to concentrate on those few features that most non-rational-choice approaches share.

2 The divide between rational choice and hermeneutic approaches is more often perceived as referring to different 'decision mechanisms' each presupposes: maximisation and rule-following, respectively. In a sense, I accept most of that line of thought in what follows (some classical criticisms to rational choice theory for not being able to deal with rule-following behaviour are Simon (1957), Sen (1970) and Vanberg (1994); see Lahno (2007) for a review); but I think that maximisation is not a deliberational mechanism that rational choice assumes people really uses, it only assumes that the behaviour of the *agents* leads to the same outcome that the *social scientists* finds as a solution to a maximisation problem, and is agnostic about the real cognitive process taking place in the 'heads' of the agents. By the way, it might perfectly well be the case that a set of psychological steps, each one consisting in a piece of rule following, might lead to the maximising solution (after all, what the social scientist does when solving the problem is to follow the rules of calculus). My suggestion is that the connection of each approach with the concept of normativity is a more essential difference between them.

3 What is more strange (to say the least), is how *little* empirical research to discover 'real' utility functions has been carried out, as compared to the effort put in building models based on 'convenient' assumptions on those functions. It is as if, in chemistry, after noting that the atomic and molecular weights of substances are essential data to be used in our formulae, almost no effort had been put into measuring the weights of *real* substances. I think the explanation of this apparent anomaly is that rational choice model builders assume that real utility functions vary a lot (not only from subject to subject, but from time to time), and they are trying to build their models with as few assumptions as possible, in order to give their models the *maximum applicability*, at the cost of a *reduced precision*.

4 Cf. for example, Gintis (2008).

5 This has some weird consequences in some branches of economics (e.g. rational expectation macroeconomics, and its derived approaches), where the assumption is often made that the agents already know the 'model' that the modeller is attempting to find! If this assumption is true, one can ask what is the added value of the modeller's work.

6 In psychology there is a parallel debate about the alternative models to the 'rationality-as-optimisation' paradigm derived from rational choice theory. Cf. Sahlin et al. (2010).

7 See, for example, Brandom (1994: 15).

8 Cf. op. cit., pp. 229 and ff.

9 Brandom's theory also demands the introduction of a different and parallel normative entity: *entitlements*, that according to him are *not* reducible to commitments by double negation (like in 'you are entitled to *X* if you are not committed to not doing *X*'), but the argument is complicated and is not necessary for the thesis I am proposing here. See Brandom (1994: 159 ff).

10 Op. cit., Chapter 4. A commitment to act is, hence, a 'reason for action' in the sense of Searle (2001).

11 Cf. Zamora Bonilla and del Corral (2008).

12 For some theories intending to do this, see, for example, Gärdenfors (1988) or Hansson (1999). For the connections between belief revision theory and economics, see Rott (2004).

13 Brandom (1994: 599 ff).

14 Brandom (1994: 26 ff).

15 For the applicability of these concepts (or some close to them) to formal models, and in particular to computer multi-agent models, see for example, Conte and Castelfranchi (2001) or Boella et al. (2008).

16 Op. cit., p. 626.

17 According to rational choice theory, it is irrational to choose actions probabilistically in such a way that, the higher the utility of an option, the higher the probability with which it is chosen; rather on the contrary, the only rational choice is deterministically identified with the option that gives the highest expected utility (because an aleatory choice will always have less utility on average than this option). But our case is completely different, for our notion of 'strength of commitment' must not be identified with 'utility' or 'expected utility'. By the way, this notion of a probabilistic connection between an agent's degree of commitment with action X and her actually performing action X serves to illuminate John Searle's notion of 'the *gap*' (Searle, 2001: 61 and ff.): the intuition that one's reasons for an action are not experienced as sufficient causes of the action itself; from the point of view defended here, the gap arises simply because the conscious process of deliberation modifies the agent's deontic score, but the action itself (including its phenomenological feeling of voluntariness) emerges out of a neural process of which the deontic score (i.e., one's degrees of commitment with this and with alternative actions) constitutes only a *partial* causal factor, the *rest* of the process being completely *unconscious* for the agent. This interpretation allows to account for the phenomenological description Searle offers of rational action, without the need of introducing metaphysical elements such as a 'substantial self' (whose own way of influencing the chain of physical processes leading to material actions remains obscure) or real gaps within the causal structure of the world (apart from the merely probabilistic indeterminacies that might depend on quantum events), and without concluding that consciousness is an epiphenomenon (since conscious deliberation has a real causal influence in modifying one's deontic score, which in turn affects unconsciously the probabilities of one's choosing an action or another).

Introduce new paragraph (full stop).

18 Cf., Carruthers (2006), Evans (2008).

19 The combination of these elements within a model will usually have the form of a *Markov chain* (cf. Ch. 13): for each possible ordered pair of 'states of the world' (as described by the commitments each agent undertakes in them, and the rest of the relevant circumstances), we will have a certain probability of passing from the former to the latter. This allows us to calculate the probability of certain outcomes (subsets of states), given any initial situation. This technique is particularly appropriate for computer modelling.

20 Of course, I don't claim that deliberation according to inferential rules is not a causal process; only that it is just a *subclass* of the whole of psychological mechanisms underlying the dynamics of our cognitive states.

21 A paradigmatic explanation of joint commitments is in Gilbert (1996).

22 See for example, Kusch (2002).

23 More literally understood, this is not exactly true, for there are cognitive mechanism that are *specific* for the interpretation of some sensory evidence as a human being, as human action, or as a social situation; but these mechanisms enter into the same cognitive equipment we have been biologically –not socially– endowed to cope with our physical and social environment.

24 I thank José Luis Ferreira for comments and criticisms on a previous version of this section.

25 Of course, an agent's valuation of a situation is not the same thing as her wellbeing in that situation (she might value other things, apart from wellbeing).

26 Cf. Osborne and Rubinstein (1994: esp. Chapters 14 and 15).

27 Actually, this is also a simplification: we should add, at the least, the probability of the player not behaving 'rationally' neither in the hermeneutic nor in the strategic sense.

28 I am conscious of the fact that 'thinking strategically' is no less 'thinking' than engaging in a collective deliberation; so, we have to take into account that the reasoning process that one player adopts when she does not expect to find out a collective agreement about what is the best option, or when she does not want to cooperate in its achievement, can nevertheless be represented as an *inferential* process, as depicted in part II, save that in this case the 'active' commitments, those more probably leading to actual behaviour, will be those based on her personal *desires*.

29 See Zamora Bonilla (2006, 2010) for the application of the model to the philosophy of science.

30 By the way, these two *different* reasons why a doxastic commitment can not be transferable to other agents allow us to explain a curious paradox in the common use of the term 'belief': on the one hand, by saying that one believes something we can mean that she *is sure* of it (it is in this sense that we talk of knowledge as a *species* of belief, for example, as justified true belief); on the other hand, we can

also mean that one *is not sure* of what he believes (as when we speak of 'degrees of belief'). The first sense of 'belief' is the one preferred in analytical epistemology, and the second one in Bayesian epistemology. It is a curious fact that almost no philosophical research has been devoted to the question of why the term 'belief' can be used to mean two *contradictory* concepts.

31 The norms governing this transfer of commitments can relate to moves from individual to individual (as in testimony), or from the members of a group to the group itself as a collective agent (as in judgment aggregation or collective belief), or vice versa.

REFERENCES

Boella, G., L. van der Torre, and H. Verhagen (eds) (2008) Special Issue on Normative Multiagent Systems, *Autonomous Agents and Multi-Agent Systems*, 17(1).

Brandom, R. (1994) *Making It Explicit*. Cambridge, MA: Harvard University Press.

Carruthers, P. (2006) *The Architecture of the Mind*. Oxford: Oxford University Press.

Conte, R. and C. Castelfranchi (2001) 'From conventions to prescriptions. Towards an integrated view of norms', *Artificial Intelligence and Law*, 7: 323–340.

Craig, E. (1990) *Knowledge and the State of Nature*. Oxford: Oxford University Press.

van Eemeren, F. and R. Grootendorst (2004) *A Systematic Theory of Argumentation: the Pragma-Dialectical Approach*. Cambridge: Cambridge University Press.

Evans, J.S.B.T. (2008) 'Dual processing accounts of reasoning', *Annual Review of Psychology*, 59: 255–278.

Gärdenfors, P. (1988) *Knowlede in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, MA: The MIT Press.

Gilbert, M. (1996) *Living Together: Rationality, Sociality, and Obligation*. Lanham, MD: Rowman and Littlefield.

Gintis, H. (2008) *The Bound of Reason: Game Theory and the Unification of Behavioural Science*. Princeton: Princeton University Press.

Habermas, J. (1981) *Theorie des kommunikativen Handelns*. Frankfrut am Mein: Suhrkamp.

Hansson, S.O. (1999) *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Dordrecht: Kluwer.

Kusch, M. (2002) *Knowledge by Agreement: The Programme of Communitarian Epistemology*. Oxford: Oxford University Press.

Lahno, B. (2007) 'Rational choice and rule following behaviour', *Rationality and Society*, 19(4): 425–450.

Osborne, M.J. and A. Rubinstein (1994) *A Course in Game Theory*. Cambridge, MA: The MIT Press.

Perelman, Ch. and L. Olbrechts-Tyteca (1958) *Traité de l'Argumentation: La Nouvelle Rhétorique*. Paris: Presses Universitaries de France.

Pettit, P. (2001) *A Theory of Freedom: From the Psychology to the Politics of Agency*. Oxford: Polity Press.

Rott, H. (2003) 'Economics and economy in the theory of belief revision', in Vincent F. Hendricks, Klaus F. Jørgensen und Stig A. Pedersen (eds) *Knowledge Contributor*. Dordrecht: Kluwer. pp. 57–86.

Sahlin, N.E., A. Wallin and J. Persson (2010) 'Decision science: from Ramsey to dual process theories', *Synthese*, 172: 129–143.

Searle, J.R. (2001) *Rationality in Action*. Cambridge, MA: The MIT Press.

Sen, A. (1970) *Collective Choice and Social Welfare*. San Francisco: Holden Day.

Simon, H.A. (1957) *Models of Man*. New York: John Wiley and Sons.

Toulmin, S. (1958) *The Uses of Argument*. Cambridge: Cambridge University Press.

Vanberg, V. (1994) *Rules and Choice in Economics*. New York: Routledge.

Walton, D. (2007) *Dialog Theory for Critical Argumentation*. Amsterdam: John Benjamins Publishers.

Zamora Bonilla, J.P. (2006) 'Science as a persuasion game: An inferentialist approach', *Episteme*, 2: 189–201.

Zamora Bonilla, J.P. (2010) 'Science: the rules of the game', *The Logic Journal of the IGPL*, 18: 294–307.

Zamora Bonilla, J.P. and M. del Corral (2008) 'Also sprach der homo oeconomicus', *Journal of Economic Methodology*, 15(3): 241–244.