4

Giacomo Bonanno

Professor

University of California, Davis, USA

---

1. Why were you initially drawn to epistemic logic?

In my case it took a long and tortuous path to reach the pasture of epistemic logic. My academic career started with a degree in Law from the university of Turin, Italy. During my (compulsory) military service, following graduation, I learned of a scholarship to pursue graduate studies in economics in England. I was lucky enough to obtain it and went to Cambridge University for a one-year Master's degree in economics. I then had the privilege of spending one year at the Mathematics Institute of the University of Warwick, doing research (in catastrophe theory) under the supervision of Christopher Zeeman. The next three years were spent in the Ph. D. program at the London School of Economics, under the supervision of Oliver Hart and John Sutton. My thesis was in theoretical Industrial Organization, which is essentially game theory applied to issues of interaction among firms. I continued to do research in Industrial Organization during the next two years as a Research Fellow at Nuffield College, Oxford. During that time, however, I became somewhat dissatisfied with the field of Industrial Organization since there was a widespread impression that, by a suitable choice of game and solution concept, almost any kind of behavior could be rationalized. In some cases the difference in results could be explained in terms of different structural assumptions about the industry (e.g. differentiated products versus homogeneous products), while in other cases the divergence of results was due to a different sequencing of moves in the game and/or the use of different solution concepts. That was the time of the "refinement of Nash equilibrium" program, which yielded as many as thirty different proposals for "the rational solution" of a non-cooperative game! I thus became interested in more foundational issues and, in particular, in pure game theory and the

conceptual underpinnings of the various solution concepts proposed in the literature. While I was at Oxford I had a long and illuminating conversation with Michael Bacharach, who encouraged me to pursue my new research interests. In 1987 I moved to the University of California, Davis. In 1994, while on sabbatical leave at Harvard university, I attended a course in modal logic in the philosophy department, taught by Charles Parsons. I clearly saw the relevance of modal logic in general, and epistemic logic in particular, to game theory. This cemented the new direction in my research, centered on the application of various branches of modal logic (epistemic, doxastic, temporal, conditional logic) to the analysis of games. While pursuing this line of inquiry I became aware of related research in other fields, notably artificial intelligence, philosophy, logic and cognitive psychology. In 1996 Mamoru Kaneko and Philippe Mongin invited me to co-organize the second LOFT (Logic and the Foundations of Game and Decision Theory) conference. I found the interdisciplinary character of the conference to be very stimulating and was happy to remain involved - together with the computer scientist Wiebe van der Hoek - in the organization of this biannual event, which is now in its ninth edition. The LOFT conferences brought to light the fact that the tools and methodology that were used in the investigation of foundational issues in game theory were closely related to those already used in other fields, notably computer science, philosophy and logic. Epistemic logic turned out to be the common language that made it possible to bring together different professional communities. It became clear that insights gained and the tools used in one field could benefit researchers in a different field. Indeed, new and active areas of research have sprung from the interdisciplinary exposure provided by the LOFT conferences.[1] I myself have benefited greatly from entering into contact with researchers from other disciplines, in particular Johan van Benthem and his collaborators and students.

[1] See: http://www.econ.ucdavis.edu/faculty/bonanno/loft.html. There is now a substantial overlap between the LOFT community and the community of researchers who are active in another biannual event, namely the TARK conferences (Theoretical Aspects of Rationality and Knowledge: www.tark.org).

2. What example(s) from your work, or work of others, illustrates the relevance of epistemic logic?

My interest in epistemic logic originated in game theory and thus I will limit my observations to the relevance of epistemic logic to game theory.

Game theory can be thought of as being composed of two separate modules. The first module consists of a formal language for the description of interactive situations, that is, situations where several individuals take actions that affect each other. This language provides alternative descriptions, from the more detailed one of extensive forms to the more condensed notions of strategic form and coalitional form. The language of game theory has proved to be useful in such diverse fields as economics, political science, military science, evolutionary biology, computer science, mathematical logic, experimental psychology, sociology and social philosophy. The unifying role of the game-theoretic language has been a major achievement in itself. The second module is represented by the collection of solution concepts. A solution concept associates with every game in a given class an outcome or set of outcomes. Most of the debate in game theory has centered on this module, in particular on the rationale for, and interpretation of, different solution concepts. The "epistemic foundation program" in non-cooperative game theory tries to determine what assumptions on the beliefs and reasoning of the players are implicit in various solution concepts. The first step in this direction was taken by Bernheim (1984) and Pearce (1984) whose aim was to identify, for every game, the strategies that might be chosen by rational and intelligent players who know the structure of the game and the preferences of their opponents and who recognize each other's rationality and reasoning abilities. In order to address this issue, one needs to answer two questions: (1) under what circumstances is a player rational? and (2) what does 'mutual recognition' of rationality mean? While there is agreement among game theorists about some essential ingredients of the notion of rationality (the most basic ingredient being: choosing a strategy which is optimal given one's beliefs), it seems that there is no clear and commonly accepted definition of rationality in general.[2] On the other hand, there does seem to be agreement that the notion of

[2] Witness, for example, the debate between Robert Aumann and Ken Binmore concerning backward induction in perfect-information games [Aumann (1996), Binmore (1996)].

'mutual recognition' of rationality is to be interpreted as 'common belief' of rationality. This is the point of entry for epistemic logic: how does one formalize the notion of common belief? What are the properties of common belief? Does common *knowledge* of rationality have different implications from common *belief* of rationality? etc. While Bernheim and Pearce captured the notion of common belief of rationality only informally, the most important contributions from the point of view of establishing a connection with epistemic logic were (implicitly) Aumann (1987) and (explicitly) Stalnaker (1994, 1996). Both authors follow a semantic approach, using structures that had been introduced in philosophy and logic in the early 1960s by Kripke (1963).[3] Aumann restricts attention to the notion of knowledge, while Stalnaker allows for the more general notion of belief. The connections between epistemic logic and the game-theoretic literature on the epistemic foundations of non-cooperative solution concepts are reviewed in detail in Battigalli and Bonanno (1999).

Other branches of modal logic have also proved to be useful in elucidating game-theoretic concepts. For instance, extensive-form games have a clear connection to the branching-time frames studied in temporal logic and thus temporal logic can shed light on solution concepts for extensive-form games [Bonanno (2001, 2002)] or on the conceptual content of properties such as perfect recall [Bonanno (2004)].

Logicians [notably, Johan van Benthem (2001, 2010)] and computer scientists [notably, Joe Halpern (2001, 2002)], have recently been very active in the study of game-theoretic concepts from the standpoint of modal logic. For a recent survey of the literature see van der Hoek and Pauly (2006).

3. What is the proper role of epistemic logic in relation to other disciplines, for instance mainstream epistemology, game theory, computer sciences or linguistics?

I once heard Robert Aumann remark at a workshop in Stanford that one important difference between game theory and pure mathematics is that game theory is about life, everyday life. I would argue that the same is true of epistemology, broadly conceived. We make decisions (some more important, some less so)

[3] A *syntactic* approach that uses the language and methods of epistemic logic to analyze games with ordinal payoffs can be found in Bonanno (2008).

on a daily basis, and we reach those decisions after having formed, consciously or unconsciously, a mental representation of the situation we face. A central part of this mental representation consists of our beliefs about the world and about the likely consequences of alternative courses of action. In a social context we also try to put ourselves in the shoes of other people in an attempt to anticipate their choices and their potential reactions to our actions. Sometimes our beliefs turn out to be correct and sometimes they are revealed to be based on misinformation or an incorrect appraisal of the situation, in which case we are prompted to revise our initial beliefs. I would argue that mental states, and in particular beliefs, are the essential element of any theory of social interaction. This is particularly true in game theory. There certainly are settings, such as a game of chess or a game of bridge, where one can refer to "the game" being played, with a common recognition among the players of the rules of the game and of the objectives of each player. However, such settings are rare. Most of the time we interact socially in situations where there are no rigid rules to be followed. In such cases Ann's mental representation of the situation might be different in some (perhaps small, but essential) elements from Bob's mental representation. "The game" being played according to Ann might be quite different from "the game" perceived by Bob, and yet each player may be certain that "the game" is "evident" or "commonly understood". Is it possible for Ann to believe that it is common belief between her and Bob that they are playing game G_1, while at the same time Bob believes that it is common belief between him and Ann that they are playing game G_2, when G_1 is a different game from G_2? In order to answer this type of questions one needs to turn to logic, in particular, epistemic logic. It turns out that such a situation can indeed arise; however, if we replace 'belief' with 'knowledge' then it cannot arise.

Given the importance of beliefs in everyday life, and social interaction in particular, it is important to have at our disposal tools that enable us to reason about beliefs, whether they are beliefs about facts or, perhaps more importantly, beliefs about the beliefs of other people. This is what I see as the important role of epistemic logic in relation to game theory. The use of epistemic logic can bring to light subtle, yet important, points. For example, the standard tool in game theory for representing interactive epistemic states is the notion of information partition [see, for example, Aumann (1987)] which embodies the assumption of cor-

rect belief, also called *knowledge*. It turns out that in this setting the notion of *common* knowledge displays the same properties as the notion of individual knowledge; in particular, if something is *not* common knowledge then this fact itself is common knowledge. Does the same hold in the case of (possibly incorrect) beliefs? If one postulates the strongest rationality properties for individual beliefs,[4] are those properties also satisfied for *common* belief? The answer turns out to be negative [Bonanno and Nehring (2000a)], with implications for the notion of a common prior [Bonanno and Nehring (1999)], which is central to the theory of games of incomplete information. In an interactive context where individuals' mental states are characterized in terms of both knowledge (those propositions of which an individual is absolutely certain) and belief (those propositions about which the individual is confident) then the notion of intersubjective consistency becomes even more subtle [Bonanno and Nehring (2000b)]. It is hard to see how one could reason about interactive mental states without the tool of epistemic logic.

4. Which topics and/or contributions should have had more attention in late 20th century epistemic logic?

It is hard for me to answer this question, since my interest in epistemic logic (and modal logic in general) was motivated by potential applications to game and decision theory. Hence my knowledge of the epistemic logic literature is somewhat limited. I am probably myself unaware of contributions to epistemic logic that might be potentially important from the point of view of game theory. I certainly believe that if game theorists had been aware, earlier on, of epistemic logic and the pioneering work of Kripke (1963), Hintikka (1962) and Lewis (1969), the literature on the epistemic foundations of non-cooperative game theory would have followed a simpler path and the theory of games of incomplete information [Harsanyi (1967-68)] might have been developed in a simpler, more transparent and more straightforward way.[5]

---

[4] That is, *consistency* (if you believe A then you do not also believe not-A), *positive introspection* (if you believe A then you believe that you believe A) and *negative introspection* (if you do not believe A then you believe that you do not believe A).

[5] Here I am echoing Robert Aumann's remark [Aumann (1999, p. 295)] that the infinite-hierarchies-of-beliefs construction is very convoluted.

5. What are the most important open problems in epistemic logic and what are the prospects for progress?

Recently I have become interested in the topic of belief revision: how should a "rational" individual incorporate new information into her existing body of beliefs? Belief revision is important in game theory, especially in dynamic games with imperfect information. If a player finds herself at an information set that was ruled out by her prior beliefs (she attached zero probability to that event), she needs to revise those beliefs (in particular, her beliefs about past moves of her opponents) by incorporating the new information. Bayes' Rule is not applicable to updating after zero probability events. Does "rationality" impose any constraints on how one should update those beliefs?

The theory of belief revision has been developed mainly by philosophers and computer scientists. The dominant approach is known as the AGM theory, following the pioneering contribution of Alchourrón, Gärdenfors and Makinson (1985). The AGM theory deals with the transition from a belief state to a new belief state in response to a piece of information. Information is treated as veridical and the "success axiom" is assumed, which requires that information be believed. While belief revision is an active area of research, there are important open problems that ought to be addressed or further explored.

The first problem concerns the notion of information. Belief revision is about incorporating reliable information into one's beliefs. What constitutes reliable information? Years ago, perhaps, a photograph could be taken as "indisputable evidence". Nowadays, with the advent of sophisticated image-editing software, photographs can be manipulated to misrepresent facts or to create the appearance of an event that did not happen. For example, in March 2004 a political advertisement for George W. Bush, as he was running for president, showed a sea of soldiers at a public event; later the Bush campaign acknowledged that the photo had been doctored, by copying and pasting several soldiers.[6] Videos and voice recordings are, nowadays, equally manipulable. What can one trust as a source of reliable information? The testimony of a witness? A newspaper article? A book? A television news report? A claim by the president of the USA? Many of us rely on the internet for information. Can material found on the internet

[6] For an interesting account of photo tampering throughout history, see http://www.cs.dartmouth.edu/farid/research/digitaltampering/.

be trusted as accurate? In Footnote 6 I gave a reference to a web page reporting photo tampering throughout history: can one be sure that the information given there is correct? Note that I am not necessarily implying that incorrect information is conveyed maliciously with the intent to mislead, although sometimes this does in fact happen.[7] It may simply be the case that an initial piece of incorrect information gets reproduced (in good faith) by different sources and thus becomes "confirmed" information. One is left wondering if, nowadays, there is *any* source of information that is completely reliable. The theory of belief revision needs to address the issue of belief formation and revision in a world where no information can be fully trusted. Furthermore, in a social context, the incentives to convey wrong information need to be studied and incorporated into a theory of belief revision.

A second problem is how to deal with sequences of items of information which are in partial or full contradiction with each other. This can happen when the same source, over time, provides contradicting information or when different sources provide conflicting information (e.g. different experts). To some extent, this issue has been studied in the literature on *iterated* belief revision, where various principles have been suggested (for example, the principle that the most recent item of information should prevail over earlier ones). However, the proposed principles seem rather *ad hoc* and in need of a firmer foundation.

A third problem concerns the notion of "minimal" belief change. The AGM theory is often referred to as a theory incorporating the principle that beliefs should be changed in a minimal way, so as to ensure that there is minimal loss of prior beliefs. While this is true

[7] Even reputable sources can sometimes be manipulated. A clear illustration of this can be found in the following newspaper report (The Sacramento Bee, September 1, 2000): "Mark J. made a big bet in mid-August that Emulex shares would decline [...] Instead they soared, leaving him with a paper loss of almost $100,000 in just a week. So J. took matters into his own hands. [...] On the evening of August 24, he sent a fake press release by e-mail to Internet Wire, a Los Angeles service where he had previously worked, warning that Emulex's chief executive had resigned and its earnings were overstated. The next morning, just as financial markets opened, Internet Wire distributed the damaging release to news organizations and Web sites. An hour later, shareholders in Emulex were $2.5 billion poorer. And J. would soon be $240,000 richer. [...] The hoax [...] was revealed within an hour of the first news report and Emulex stock recovered the same day. Still, investors who [believing the fake news release] panicked and sold their shares, or had sell orders automatically executed at present prices, are unlikely to recover their losses"

when new information is compatible with prior beliefs, in the case where the new information contradicts the earlier beliefs, there is really no constraint imposed by the AGM postulates in terms of preserving as many of the old beliefs as possible. Indeed one way of revising beliefs, which is consistent with the AGM postulates, if to form a new belief set consisting exclusively of the learned information and anything that can be logically deduced from it. More work needs to be done on what minimal belief change entails.

A fourth problem is related to the observation that, while epistemic logic deals with how people *should* reason, how they *should* form beliefs and how they *should* respond to new information, an important area of research concerns how people *actually* reason and process information [Holyoak and Morrison (2005), Johnson-Laird (2006)]. It seems to me that much would be gained by attempting to integrate insights from both areas of research.

Finally, while beliefs are arguably the most important mental states, there are other mental states that deserve a similar in-depth investigation, in particular intention, prediction and emotions and their interactions with belief formation and belief change.

Although some contributions have addressed the "open" problems listed above, more work needs to be done in order to gain a better understanding of the issues involved. Furthermore, interdisciplinary approaches might be useful in shedding light on connections between seemingly different research questions.[8]

References

Alchourrón, Carlos, Peter Gärdenfors and David Makinson (1985), On the logic of theory change: partial meet contraction and revision functions, *The Journal of Symbolic Logic*, 50: 510-530.

Aumann, Robert (1987), Correlated equilibrium as an expression of Bayesian rationality, *Econometrica*, 55: 1-18.

Aumann, Robert (1996). Reply to Binmore. *Games and Economic Behavior,* 17: 138–146.

Aumann, Robert (1999), Interactive epistemology I: Knowledge, *International Journal of Game Theory*, 28: 263-300.

[8] For example, it turns out that the mathematical structures used in decision theory to model revealed preference can be re-interpreted in terms of partial belief revision in accordance to the AGM postulates [Bonanno (2009)].

Battigalli, Pierpaolo and Giacomo Bonanno (1999), Recent results on belief, knowledge and the epistemic foundations of game theory, *Research in Economics*, 53: 149-225.

Bernheim, Douglas (1984), Rationalizable strategic behavior, *Econometrica*, 52: 1007-28.

van Benthem, Johan (2001), Games in dynamic epistemic logic, *Bulletin of Economic Research*, 53: 219–248.

van Benthem, Johan (2010), *Logic in games*, Amsterdam University Press, Texts in Logic and Games, Amsterdam (to appear).

Binmore, Ken (1996). A note on backward induction. *Games and Economic Behavior,* 17: 135–137.

Bonanno, Giacomo (2001), Branching time logic, perfect information games and backward induction, *Games and Economic Behavior*, 36 : 57-73.

Bonanno, Giacomo (2002), Modal logic and game theory: two alternative approaches, *Risk Decision and Policy*, 7: 309-324.

Bonanno, Giacomo (2004), Memory and perfect recall in extensive games, *Games and Economic Behavior*, 47: 237-256.

Bonanno, Giacomo (2008), A syntactic approach to rationality in games with ordinal payoffs, in: Giacomo Bonanno, Wiebe van der Hoek and Michael Wooldridge (editors), *Logic and the Foundations of Game and Decision Theory*, Texts in Logic and Games Series, Amsterdam University Press, 59-86.

Bonanno, Giacomo (2009), Rational choice and AGM belief revision, *Artificial Intelligence*, 173: 1194-1203.

Bonanno, Giacomo and Klaus Nehring (1999), How to make sense of the common prior assumption under incomplete information, *International Journal of Game Theory*, 28: 409-434.

Bonanno, Giacomo and Klaus Nehring (2000*a*), Common belief with the logic of individual belief, *Mathematical Logic Quarterly*, 46: 49-52.

Bonanno, Giacomo and Klaus Nehring (2000*b*), Intersubjective consistency of Knowledge and Belief in: Martina Faller, Stefan Kaufmann and Marc Pauly (editors), *Formalizing the dynamics of information*, CSLI Publications, Stanford, 27-50.

Halpern, Joseph (2001), Substantive rationality and backward induction, *Games and Economic Behavior*, 37: 425-435.

Halpern, Joseph (2002), Characterizing the common prior assumption, *Journal of Economic Theory,* 106: 316-355.

Harsanyi, John (1967-68), Games with incomplete information played by "Bayesian players", Parts I-III, *Management Science*, 8: 159-182, 320-334, 486-502.

Hintikka, Jaakko (1962), *Knowledge and belief*, Cornell University Press, 1962.

van der Hoek, Wiebe and Marc Pauly (2006), Modal logic for games and information, in: Johan van Benthem, Patrick Blackburn, and Frank Wolter (editors), *Handbook of Modal Logic*, Elsevier, 1077-1148.

Holyoak, Keith J. and Robert G. Morrison (editors) (2005), *The Cambridge handbook of thinking and reasoning*, Cambridge University Press.

Johnson-Laird, Philip (2006), *How we reason*, Oxford University Press.

Kripke, Samuel (1963), A semantical analysis of modal logic I: normal propositional calculi, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9: 67-96.

Lewis, David (1969), *Convention: a philosophical study*, Harvard University Press.

Pearce, David (1984), Rationalizable strategic behavior and the problem of perfection, *Econometrica*, 52: 1029-50.

Stalnaker, Robert (1994), On the evaluation of solution concepts, *Theory and Decision*, 37: 49-74.

Stalnaker, Robert (1996), Knowledge, belief and counterfactual reasoning in games, *Economics and Philosophy*, 12: 133-163.