1

Johan van Benthem

University Professor and Professor of Logic

University of Amsterdam, The Netherlands

Stanford University, USA

Why were you initially drawn to game theory?

I should say at the start that I am not a game theorist. I am a logician enamoured of game theory – of course, in a purely Platonic sense – and accordingly, I tend to idealize the object of my affections. It would be tedious for me to pronounce on questions like 'where game theory should go': although, wherever, I do like the way it walks ...

In my early student days in the late 1960s, revolution was in the air, and we did not take it for granted that what our professors (addressed by their first names from the start, and shivering every time we quoted from half-understood revolutionary German texts) told us to read was the real stuff. That is how physics students like me found ourselves taking courses in abstract mathematics – I first heard about Category Theory through the student grapevine –, or even on the stairs of the Humanities building on our way to classes in generative grammar, or the history of Vietnam. I remember coming home one day and telling my old landlady that I had just learnt that one could prove mathematically that the Dutch language had infinitely many sentences. She looked at me strangely, and said "Johan, don't be silly". That was that. In a science faculty starved of female students, those visits of course also expressed our cravings for fashion, beauty and elegance. Eventually, those excursions took me to a class on logic, and I have been hooked ever since, switching to mathematics and philosophy. But none of these courses we took actually referred to game theory. *That* was rather a subject for books which I read in addition to our standard fare, such as Kemeny, Snell & Thompson's *Finite Mathematics* which had a tiny bit of matrix games and solutions in mixed strategies. Now the nice thing with the topology of the academic literature is that it is such a highly interconnected 'Cultural Park'. You can enter anywhere, and then there are all these wonderful trails that soon take you to unexpected new landscapes. I quickly found Luce & Raifa's book *Games and Decisions*, which conveyed the excitement of new things going on, and the surprises of mathematical structure in what look like garden-variety human interactions. There was even the secret thrill of forbidden fruits, since someone had told me that some radicals in Sweden had proposed replacing logic by game theory as the formal apparatus which every philosopher should know. Why not at least invest a bit in this possible future world, whether or not it materialized?

A second appeal came from the cheap and immensely informative German paperback series called 'Hochschultaschenbücher' (University Pocketbooks) published in Mannheim, which contained didactic master pieces by major German professors, often even uncompromising original scientific contributions, at virtually no cost at all. These publishers were benefactors of humanity, and I am still grateful to what they did for us. One of the books I bought was Paul Lorenzen's *Logische Propädeutik*, his wonderful little monograph seeking the foundations of logic neither in the intimidating austerity of mathematical proof, nor in the lush realities of semantic truth, but rather in the structure of successful dialogical interaction. According to Lorenzen, valid arguments are those patterns from premises to conclusions in which the proponent of the conclusion has a winning strategy against any opponent granting the premises. Thus, there is a third independent pragmatic intuition of logical validity, based on viewing argumentation as a game. I have been converted to that view ever since, even though most of my professional life has been under camouflage as a model theorist, or occasionally a proof theorist. Additional evidence for Lorenzen's view came from the way in which logical operations find their natural place as operators of dialogue control in the game-theoretic setting: conjunctions and disjunctions are choices by the two players, negations are role switches. Again, this dynamic intuition just rings true to me. Nowadays, in my later years, I would also see this historically as a way of rethinking the past. Many people think that the origins of logic in Greek Antiquity come from mathematics, with Euclid's *Elements* as the paradigm of deductive proof, or the empirical sciences, with Aristotle's syllogisms as the engine of classification. But what may be more likely is that these origins lie in the debating practices of the Greek polis, for which the

Sophists trained their students, with their high-brow expression in the dialogical format of Plato's *Dialogues*. And similar issues exist elsewhere: I just heard a talk on Indian logic suggesting that its origins might lie in legal practice, i.e., again, dialogue and debate.

Even so, my love for game theory did not extend beyond teaching Lorenzen games, and later also Hintikka's evaluation games and Ehrenfeucht model comparison games, to my students – both philosophical and mathematical – as a supplement to the usual way of pouring the basics of logic into young adolescents. I did keep an eye open toward uses of games in logic and surrounding areas like the philosophy of science (Robin Giles' operational semantics for physics comes to mind), but left it to more ideological proponents of game methods in logic and argumentation theory to hold torches and make speeches. My only published effort is a little survey paper on 'Games in Logic' from 1987, in which I listed all uses of game theory in logic which I knew for the German Fraunhofer Foundation – for the sum of 1500 Deutschmarks, then a considerable amount of money for a Dutch professor. Of course, games entered my world now and then. For instance, when editing the *Handbook of Logic and Language* with Alice ter Meulen, the eventual version in 1997 lists the Hintikka–Sandu game-theoretical semantics of meaning in terms of imperfect information games in the Top Five of major paradigms in understanding natural language. But game-theoretical semantics is just a weak reflection of the richness of actual game theory, and it does not formulate a broader program. (I did write a paper in the early 1990s for a Hintikka celebration playing with a 'Church Thesis for Games', saying that, just as every computation can be mimicked by a Turing machine, every rational interaction is playable as a game.) More significant contacts between the broader epistemic logic community and game theorists had already been pioneered by then by Joe Halpern in the TARK community, by Robert Stalnaker in his work over the 1990s, and by Wiebe van der Hoek and Giacomo Bonanno in the starting of the LOFT conference series. But still, I just kept observing.

Things heated up considerably during the time of my Spinoza Award Project *Logic in Action*, when Paul Dekker and Yde Venema organized a workshop on games in 1998, where, for the first time in the history of our Institute of Logic, Language and Computation, we had a game theorist as an invited speaker, namely Arnis Vilks from Leipzig. We saw at once how congenial all this was, and how logicians and game theorists are really brothers-in-arms, or at the very least, cousins-in-arms. This also demonstrated that mutual flow of ideas was possible: from game theory into logic, as before, but also *from logic into game theory*! I started teaching a graduate seminar on 'Logic and Games' at Amsterdam and Stanford, which has been running essentially until today, resulting in various dissertations on the border line of logic, game theory, and computer science. Several of them, by Marc Pauly, Boudewijn de Bruin, and Merlijn Sevenster have already attracted quite some attention. And other talents have emerged at this fault-line between disciplines elsewhere, such as Paul Harrenstein, Sieuwert van Otterloo, or Francien Dechesne. Ever since those days, we have had a continuing series of encounters in The Netherlands, where games became a popular theme in many places. For instance, this year, we will have the 15th instalment of our informal but high-powered workshops on Logic, Games and Computation. And also at ILLC, we suddenly find that games are a unifying interest among our leading linguists, mathematicians, and computer scientists, not for profit, but for insight, and for fun! As a reflection of all this, the ILLC obtained its European Marie Curie Centre 'Gloriclass' bringing together some 15 Ph.D. students at the interfaces between all these disciplines, while also creeping up on cognitive science now and then. By now, we also hope to take this style of thinking to a European scale, in the project 'LogiCCC' of the European Science Foundation on logics for intelligent interaction, where we are joining with like-minded people on related interfaces all across Academia. A personal research interest in games and multi-agent interaction naturally leads, at least to me, to a desire for social community building!

What example(s) from your work (or the work of others) illustrates the use of game theory for foundational studies and/or applications?

As I said, I see two directions to the contact between logic and game theory. Let's first take the one *from game theory to logic.* I have already indicated how even just basic ideas from game theory seem congenial to notions at the very heart of logic. Many people think that interaction is just some 'nuisance' for true logic, arising from the - perhaps unfortunate - fact that we populate this planet simultaneously with many others, resulting in tons of gossip, quarrels, and cowardly compromises. (Recall that the famous Dutch logician and mathematician Brouwer was a solipsist: he heroically decided to ignore this social feature altogether.) I think, by contrast that interaction, and the resulting ‘Many Mind Problems’, are just as central to logic as ‘Many Body Problems’ are to any significant physics. And game theory has provided notions which make this feeling precise, tying in the absolutely basic notions of truth, proof, or invariance between models with strategies in multi-player games representing different roles in enquiry. These roles range from Verifiers versus Falsifiers, Proponents versus Opponents, Duplicators versus Spoilers, or Builders versus Destroyers (this is beginning to sound like Hindu theology, but so be it). In the hands of distinguished logicians like Lorenzen, Ehrenfeucht, Hintikka, Blass, Hodges, Girard, Abramsky, Väänänen, and many others, these ideas have become powerful tools for formulating logical notions, and proving their properties. This may not be common knowledge in logic textbooks or among philosophers of logic yet, but it will. But this achievement can be appreciated in two ways. One is just as a tool, perhaps even just a metaphor. This ‘Weak Thesis’ is all-right: games are both tools and metaphors. But the other, more radical view is the ‘Strong Thesis’ that games represent something essential about logical notions, and that the two fields live in pre-established harmony. That view happens to be mine.

What new insights do we get from reformulating things this way? I will mention one, and it illustrates a non-trivial issue at the same time. Consider the major paradigm of a successful logical system, first-order predicate logic. It is replete with game imagery, once you see it properly, with Abelards and Eloises behind every tree and shrub – but let’s focus on one aspect. When you view first-order formulas as denoting evaluation games, the absolutely basic issue of logical equivalence – which determines what we mean by formulas expressing ‘the same proposition’ - translates into the issue *when two games are the same*. Now there is no canonical answer to this, just as mathematics has no canonical answer to when two geometrical spaces are the same, or computer science to the question when two processes are the same. It all depends on natural notions of structure-preserving transformations and invariance. But at least, game theory suggests that we can look at various identification levels: global strategic forms, powers of players, or extensive forms. Correspondingly, we now get a finer view of equivalence levels for logical propositions, and we enter one of the most basic and vexing areas in the philosophy of logic. I have shown in a 2003 paper that on some views of game equivalence, predicate logic with the corresponding notion of propositional equivalence becomes *decidable*, pace Gödel, Turing, and Church. So, much can be at stake in getting clear on these matters!

Other benefits of this finer grain in logical structure are richer views of what logical constants are about. I would say that, viewed as interactive processes, games split standard logical notions into a great variety of natural notions of control. Take logical conjunction. One reading makes it a *choice* of sub-games for your opposing player, another the sequential composition of first playing one game and then the other, and a third natural reading makes it some sort of *parallel composition* of playing two activities either simultaneously, or interleaved. All this is highly congenial to the move in computer science from single Turing machines to distributed networks of computing agents, who involve in finite or infinite interactions, who by now are endowed with capacities for observation, message passing, and even goals and desires. I will not elaborate much on this computational process connection, but it is certainly another major strand in the total fabric of contacts that I am describing. We may not have the totally crystallized definite view of the natural repertoire of game equivalences, and matching logical constants here (with apologies to those who think they *have* given the world just that...), but, praise be to the game perspective, what a much richer conceptual world for a logician to live in!

Now here is the standard objection to all this. It may be games, but is it *game theory*? After all, real game theory is about agents who have preferences and goals, who attach values to outcomes – and its major mathematical results are about appropriate notions of strategic equilibrium, and when we can have them. Indeed, that mathematical theory revolves around mixed strategies and probabilistic considerations which may look alien to pure logic. I agree that most of this structure has not found its way into logic yet, with a few exceptions here and there, in evaluation games with imperfect information, and some process theories that allow for preferences between transitions of a system. But I see no reason at all why these perspectives could not be brought in. For instance, goals and preferences become essential once you try to understand the drift of real argumentation, winning debates, or just dispensing procedural justice as a chairman. I see these phenomena as major challenges to logic, since we want to interface our accounts of validity with rational ways for making our views prevail, or: for standing refuted when we should be. I have write some papers on 'winning debates' where a mixture of standard logic and real games is of the essence. Likewise, I have argued in print that mixed strategies in evaluation games make perfect sense as probabilistic mixtures of Skolem functions once you make the move from a deterministic to a probabilistic universe of objects, as happens in quantum mechanics.

Indeed, once you take this view, it will crop up in other places, too. Take, not argumentation, but the much-studied phenomenon of *belief revision.* Even though this is usually cast as a single agent recording incoming information and adjusting beliefs, in reality, it is first and foremost a multi-agent phenomenon, where we have to merge information from different sources, and where interactions with other minds make us change ours. Clearly, the eventual theory of belief revision must be about revising our goals as well, and again, values, preferences, and longer-term strategic interaction will be of the essence. This point was also made in the context of learning theory by Kevin Kelly, who shows that only in this way, can one compare different revision strategies as to their success toward stated goals. Actually, it seems to me that the *linguists* are ahead of the logicians here right now. In the work of Lewis, Parikh, Jaeger, van Rooij, Gärdenfors, and others, the crucial function of language is communication, and stable meanings emerge as equilibria in formal, but somewhat realistic, coordination games.

Now, let's look the other way, and go from *logic to game theory.* Here some of the earlier themes return, but now with a reverse thrust. And some new ones get added, since we are now looking at general games with the tools of logic. But of course, the real situation is that of a meeting of disciplines with ideas flowing both ways. For instance, the study of games is a natural continuation of the study of *process structure*, which I see as one of the major 'cultural' contributions which computer science has made to the academic landscape. Thus, game theory is giving us ideas about interactive multi-agent processes, taking it to 'the next level'. But of course, one has to merge the respective insights and modus operandi. Game theory is mainly about global notions like strategic equilibrium in a game, and it has been amazingly successful in getting away with this high-level abstraction, and extracting useful and insightful information from it. Indeed, in early stages, it even seemed as if these notions, and techniques for finding them like Backward Induction or more sophisticated fixed-point theorems, were writ in stone. Right now, I would say that this global representation needs *fine-structure*, of the sort than can be provided by ideas from both computational and philosophical logic.

The 'dowry' from *computational logic* is its sophisticated thinking in terms of process equivalences, such a bisimulation, and their matching logical languages: modal, first-order, other, describing the corresponding invariant properties of games. I have developed these themes in my 2003 JoLLI paper 'Extensive Games as Process Models' pointing out to which extent modal and dynamic logics can then provide an explicit fine-structured account of games, and very importantly, of players' strategies: perhaps the real, but somewhat unsung, heroes of game theory. I think that we are the threshold of a merge between game-theoretic equilibrium mathematics and process logics and temporal logics. In this way, we will develop a greater sensitivity to the *Balance* between expressive power and computational complexity of essential properties of games, a balance which permeates so much of computational logic.

This mix and this balance become even more delicate, when we add players with limited powers of observation: just like us. In that case, we also need the dowry of *philosophical logic*, and its accounts of knowledge, belief, and other relevant informational states. To some, this seems like a strange and unhappy mixture. Computational logic is hard-core, and almost as respectable as straight mathematical logic in the foundations of mathematics. By contrast, philosophical logic is about attitudes of fallible agents with notions referring to their individual idiosyncracies: in short, the world of imperfection, compromise, and often mere mathematical bubbles. Even so, this *is* the world of intelligent agents, and we had better use all available tools to understand them. What we are seeing now is the emergence of all sorts of theories which merge ideas from computational and philosophical logic. My own area of *dynamic-epistemic logic* is a typical example. The work there on actions of information change, belief revision, and even preference change, seems to fit seamlessly with the study of games. In my 2002 paper 'Games in Dynamic-Epistemic Logic' I give several examples of this, showing e.g., how uniform strategies are exactly the ones definable by means of 'knowledge programs'. I elaborate another strand in my 2004 paper 'Rational Dynamics', providing a new epistemic take on game-theoretic 'solution algorithms', taking them seriously as processes of inner deliberation and knowledge update. My eventual hope would be that, in this way, by operating on such a broader front, we can also systematize the game theory of solution concepts and their epistemic characterizations, which has evolved since Auman's pioneering work by the great game theorists of the 1980s. At present, it consists largely of a haphazard collection of, admittedly famous, notions and results.

All this is mainly still cooking right now, but I find this whole research area liberating. For instance, at the moment, I find myself working with my students on logics of *preference*, long considered a stagnant malarial backwater of logic. But now, we are taking the richer perspective suggested by games, inspired by the problems of Backward Induction, not as needing a 'quick fix' once and for all, but as a starting point for an in-depth dynamic analysis of preferences. We ask what leads to the preferences that we have, and how they might change dynamically under pressure of suggestions, commands, or observations of merits of other players. Van Benthem, van Otterloo & Roy 2005 gives a first example – but much more is to come.

Once again, some logicians thinks this is 'dirty' or at least 'messy'. Game theory imports *economics*, and hence thinking in terms of cost, value, and so on. By contrast, I think the latter is an essential and general intellectual perspective with its own intuitions and reasoning styles, which works across academia, enriching (excusez le mot) other disciplines which it touches.

What is the proper role of game theory in relation to other disciplines?

This question is not for me to answer. I even find it sounds too much like those old German discussions of the proper place of disciplines in some grand intellectual order of things in the late 19th century. (It is usually the working classes which need to understand their 'proper roles': the rich are free.) I think disciplines should thrive and influence other disciplines, by caring as little as possible about academic hierarchies, or who is supposed to be the guardian of what. Indeed, if I wére to say anything more – which I will now proceed to do –, I find *logic* and *game theory* very similar in their academic roles. Both provide very general models for analyzing intelligent behaviour, though focussing on different levels so far: logic by and large more micro, game theory more macro. For both, it is hard to say to which extent they are normative or descriptive (maybe that distinction has become tedious anyway), and their relationship to experimental cognitive science is delightfully tortuous. And finally, both do not just analyze given behaviour, they also provide for design of new styles of behaviour that can be incorporated into our human repertoire. They would really make a good match when paired as academic disciplines—but maybe, I already said that?

What do you consider the most neglected topics and/or contributions in late 20th century game theory?

Again, this is not for me to say, and also, I do not like the term 'neglect'. Admit to it, and before you know it, some American lawyer has sued you. Of course, there are areas where I would just like to 'hear more' from game theorists. This would be in particular in *explicit theories of strategies*, richer than what we usually get, richer accounts of the *step by step dynamics* of extensive games, rather than pre-encoding everything right at the start in some huge 'type space', and finally, instead of coming up with different games for different occasions, some systematic account of *how games can change*, and what then happens to their properties, without having to re-compute everything from scratch every time.

What are the most important open problems in game theory and what are the prospects for progress?

Here, I will just list the interfaces and developments which I expect to happen. I am seeing an emergent logic-game theoretical paradigm where logic provides the fine-structure behind the usual games, which will integrate ideas from three sources: (a) dynamic epistemic logics describing single steps of information update, belief revision, and other basic events, (b) process logics from computer science describing compositional process structure and longer-term behaviour, and (c) game-theoretic structure having to do with preference-based equilibria.

In this coming together, I expect specific teaming up between, e.g., game-theoretic fixed-point theory and fixed-point logics in computer science. Likewise, I expect exciting merges in methodology. Some people think that computational logic sits badly with real games, because its compositional methodology founders on the latter having no good notion of sub-game. This seems premature to me, since the main challenges to compositional methodology have never been easy, but the results have always been rewarding. Then, there will be a growing interface between the mathematics of dynamical systems underlying evolutionary game theory, and logics for infinite processes, perhaps even co-algebra and its modal logics as being developed today.

Note that I have cast none of this anywhere as logic being 'applied' to game theory, or game theory being applied to logic. I think those terms mean very little in significant interactions between healthy disciplines. They do not meet in order to cure each other's ailments. They meet to produce to *new offspring*, and the quality of that offspring is the test of their match.

So much for technical perspectives from game theory, logic, and computer science merging into one apparatus. But there is also the arena where these frameworks will play. Based on tell-tale signs in the avant-garde literature, I expect major influences in philosophy, ranging from interactive epistemology to the philosophy of action and social philosophy. In fact, this is a safe prediction, since so much is already going on! Likewise, as linguistics is making its interactive turn, optimality theory is transforming into game theory, and again the resulting paradigm will have greater power than either component. Finally, I see all this moving into experimental cognitive science. It is very interesting to see that logic and game theory have started picking up cognitive interests around the same time in the 1990s. As cognitive scientists will see more and more that intelligent interaction is the key to understanding human rationality and success, the logic game theory connection will become stronger accordingly – and we will be scanning 'games instead of brains'.

Was this what I saw vaguely as a student reading game theory books with my pocket light under the blankets? No. But it is what I am wishing for today, and as game theory tells us, wishes can come true, provided we play our cards right.