SOME THOUGHTS ON THE LOGIC OF STRATEGIES

Johan van Benthem

This little piece is an extended version of some notes for a wrap-up session of the Workshop 'Modeling Strategic Reasoning' at the Lorentz Center in Leiden on February 24th 2012. Its purpose is merely to collect some general thoughts on strategies triggered by our meeting, high-lighting some general issues that seemed to resonate at the Workshop.

What was the topic? For a start, take the title of our Workshop. Was our topic strategies themselves, reasoning about strategies, or modeling reasoning about strategies: two layers on top of the phenomenon itself? My view is that our meeting was about strategies really, while reasoning just happens to be the main thing that our community does in every area. But even the core meaning of the word 'strategy' turned out to be a contentious issue at our meeting. Some people thought of strategic behavior as typically involving some *structured plan* for the longer term, whereas others thought of it as essentially interest- or preference-driven behavior with *ulterior goals*, sometimes seen as high rationality, sometimes as self-serving and not really nice. My own inclination in the area has been to focus on the former aspect of strategies, and I will mainly do so here, but the two meanings are not in conflict.

Finding the key notions and core language Still on the issue of delimiting our topic, in serious conceptual analysis, we want to find core families of notions that naturally belong together. Thus, in epistemology, it makes little sense to study knowledge without at the same time studying belief, and perhaps also the notion of evidence. They form a natural trinity. Likewise, here, is strategy really the right notion to focus on, or should it be generalazed? And can it be studied on its own, or can we only get a proper perspective by adding further notions from the start? As it happens, the terminology used in our academic community shows a great variety. People work on 'strategies', 'tactics', 'plans', 'protocols', 'methods', 'agent types', and the like, and often they amount to very similar things. Is a 'Liar' a type of person, a program producing a certain behavior, a method for dealing with other people, or a strategy? Clearly, these terms are not all formally well defined, though some clues for their use do occur in natural language. In daily discourse, tactics means strategy writ small (a 'strategette'), while strategy is tactics writ large – and one feels these are

similar notions of modus operandi, but operating at different levels of describing activities. It might be worth trying to achieve some further conceptual clarification, and reserve some terms for different uses. For instance, in my own line of research, the following distinctions make sense. A *protocol* is a general style of behavior over a longer period of time, say, a persistent cooperative Gricean linguistic practice. Such a broad practice is like an operating system that still admits of running very different *strategies* inside for more particular short-term purposes, such as a way of conveying a message, or of convincing your audience of some claim. Moreover, these strategies are used by *agents*, embodied devices that can run different strategies, and conform to protocols. But I admit that I often switch between these terms, as much as other colleagues at our workshop. And I freely use even further words, such as *plan*, and it would be over-optimistic to think that this is well-defined conscious terminology. Indeed, plans and strategies are close, though I view plans as less deterministic than strategies, and also, as something one is aware of and commits to, more than strategies. Even if this terminological diversity reflects a lack of conceptual clarity, I do not think that plunging into linguistic semantics or philosophy is going to resolve these issues for us. [1] We just may have to make up our mind about the core notions, and their optimal matching terminology at some stage, but for the moment, allowing a rich notion of 'strategy' may be the best way to go, while remaining aware of the potential of miscommunicating about.

Starting from best practice Maybe at the opposite extreme of worrying about first achieving conceptual structure and clean terminology is the inductive approach of best practice. Consider the question of what is an algorithm? Maybe the best way of approaching this issue is by giving a set of concrete examples, exemplars that we see as characteristic of the notion, and that help nurture our intuitions. Think of famous sorting algorithms, geometrical constructions, or algebraic procedures. Can we do something similar for strategies? Is there a core range of evergreens that attract our attention, and help convey to a stranger what it is that exercises us so much? At the Lorentz Workshop, in one special cognitive channel, I tried to listen to talks just for examples that people found particularly motivating.

[1] This may be different for epistemology and informational action, where I do think that natural language has a rich and telling repertoire of expressions like 'know', 'suspect', 'learn', 'note', 'discover', 'tell' … that we seem to use with a certain amount of stability and even sophistication.

Frankly speaking, I did not collect a large set, and the ones I saw were somewhat abstract and logical in nature – perhaps due to a filter inherent to logicians' channels. Elegant items that struck me as giving large effects for small effort were surprisingly simple *copying strategies*: copy-cat in linear game semantics, Tit-for-Tat in evolutionary games, or 'strategy stealing' in classical game theory. What such scenarios have in common is a simple way of copying useful moves from somewhere else. At our workshop, this made me aware of new strategies of this kind. For instance, the most sophisticated examples of signaling strategies in *IF* logic were not concrete Skolem functions, the usual paradigm in that field, but generic identifying strategies like 'take the same object', 'take some other object'. Moreover, much can be built from this modest repertoire once we allow further constructions. Game semantics for linear logic builds all its strategies in this way, suggesting that understanding copying might be a major part of understanding the working of strategies. Maybe our best paradigm is *Judo*, using our opponent's moves to win it all.

Core logic of strategies Here is yet another take on the field. Logicians have their own approach to zooming in on a notion, no matter what stable terminology or exemplars trail along with it. We think that a notion is stable if there are some basic reasoning principles at its core, shedding light on a broad range of uses. I myself indeed like to believe that there is a core calculus of reasoning about strategic patterns in behavior, from first-person deliberation to third-person assessment of strategic action in games. One obvious candidate has long existed for this purpose: *propositional dynamic logic PDL* of programs with sequential operations. The compositional structure of strategies often has its typical *IF THEN* and *WHILE DO* character, and also, *PDL's* generalizing from strategies as functions on players' turns to arbitrary binary relations makes a lot of sense when we generalize to plans. Intuitively, a plan restricts my choices in helpful ways, but it need not fix my behavior uniquely. What *PDL* delivers is a general way of reasoning about generic strategies, though it can also handle specific functions or relations in single games. But is it enough?

One way of testing this is quasi-empirical, starting by cataloguing results in game theory, computer science, and logic, toward a repertoire of ubiquitous proofs concerning strategies. What is a best repertoire of paradigmatic arguments about strategies that can be mined for logical patterns? My own favorite sources are standard proofs of major results, such as the

Gale-Stewart theorem, or the Von Neumann-Morgenstern-Nash fixed-point theorem – and one should probably add more recent proofs from computer science, for theorems about memory-free strategies in automata theory, or strategies in the μ–calculus. If you look at the details of what happens in such mathematical arguments, you will see a rich amount of logical finesse that generalizes far beyond these results themselves. Of course, this still does not tell us what framework would be best for a base calculus of strategies. Propositional dynamic logic is one option, linear logic game semantics another, but one may also want to consider alternative paradigms like co-algebra with strategies in infinite games. [2]

Plans across changing situations So far, I have emphasized one key aspect of a logical understanding of strategies, a grasp of fundamental valid patterns in reasoning about them. But there are further relevant intuitions that invite a logical angle. What is important to me is that a good strategy, or more generally, a plan is something should still work when circumstances change: its should be robust under changes, at least, under small changes. But then we run into a well-known problem that led to lively discussion at our Workshop. Many programs or strategies seem very local. They just fall apart when you make small changes in a game. You can see this easily in computing the Backward Induction strategy: its optimal path can shift wildly with addition or deletion of moves. Likewise, programs meeting certain specifications may just stop working under slight changes in the model. [3]

Two options seem to arise for coping at this stage – that might be called *recomputation* versus *repair*. Should we just create a new strategy or plan in a new game, or gently revise

---

[2] Coalgebraic strategies are typically top-down objects that can be used by making an observation of their head after which an infinite tail remains. This is very different from the bottom-up behavior of terminating programs highlighted in *PDL*. I will not pursue these various approaches here, but refer to van Benthem 2012 for further discussion of calculi of strategies with associated game constructions. That book also discusses how strategies change our view of logic itself when we move from logic *of* games to logic *as* games, reading formulas themselves as complex game expressions.

[3] One might say that the flexibility we are seeking is already given in the standard notion of a game, where the strategy has to work under any eventuality. And in principle, one could then collect all relevant cases of change that I mentioned into one 'supergame', asking for a strategy working there. But the latter form of pre-encoding seems far removed from our ordinary understanding of plans.

a given plan? Looking at practice, we often seem to start with the latter scenario, and only go the former when forced by circumstances. But what would be a serious theory of plan revision? For instance, along what comparison order of plans would that take place? And can we say more precisely when gradual changes are sufficient, and when they have to be drastic? I think all this raises interesting issues of plan structure, definability and model-theoretic preservation behavior that we have not yet begun to address systematically. [4]

Plans, knowledge and understanding My discussion so far fits a conception of strategies as pure algorithms composed out of atomic actions and factual tests. But in most topics studied at our Workshop, pure action was not enough. *Information* plays a crucial role in plans and strategic behavior by the sort of agents that we study. This topic is much studied, and we know a lot about how to extend the above approaches. In particular, in *PDL* style one can get a long way with epistemic dynamic logics for planning (Moore 1985) or knowledge programs (Fagin, Halpern, Moses & Vardi 1995). [5] There is no need to discuss this line of work here, even though it is highly relevant to logic of strategies in my sense.

Instead, I just want to point out that, despite this success, basic questions remain about the entanglement of information and action. In addition to plans involving knowledge, there is also the fundamental notion of 'knowing a plan' that seems crucial to rational agency. There is no generally accepted current explication of what this means. One common line is to ask for a sufficient amount of propositional knowledge about what effects the plan, and remaining parts of it, will achieve. [6] But intuitively, more is involved in knowing a plan. Consider what we want genuine *learning* to achieve: not just correct propositional knowledge, but also the ability to engage in a certain practice based on the plan. In education, we teach *know-how* at least as much as 'know-that'. How can we add the latter notion to the

---

[4] Compare the nice example of repairing programs discussed in Huth & Ryan 2004. We know very little by way of systematic results. Thus, I am not even aware of model-theoretic preservation theorem under submodels or model extensions for such a simple logic as *PDL* with programs.

[5] Such epistemic extensions do not seem to exist yet for other formats, like linear game semantics.

[6] This issue plays very concretely in the area of 'epistemic planning' (Birkegaard Andersen et al. 2012), where different kinds of knowledge or beliefs become important: about where we are in following some current plan, but also beliefs about how we expect the process to develop over time.

formal logics of propositional knowledge that we have developed so far? This contrast may be highlighted in terms of *understanding* a strategy or a plan versus merely 'knowing' it. What is the intuitive surplus of understanding over knowledge? This issue resonated at the Workshop, and it triggered much interesting discussion. In addition to propositional knowledge of what a plan, or parts of it, actually achieves as it is being followed, people mentioned more modal desirable features such as the earlier robustness: counterfactually knowing the effects of a plan under changed circumstances, or the ability to modify it as needed. [7] Other desiderata included being able to describe a plan at different levels of detail, moving up or down between grain levels as needed. I am certain that there are more key aspects, but my purpose here is just to raise the issue for the reader's consideration. [8]

Entanglement with preferences and goals Having looked at combining action and information, let us now briefly consider toward the other sense of 'strategic' behavior in our introduction, that of being based on motives and golas. As a concrete instance, consider how action and *preference* are entangled in game theory and practical reasoning in general. Many of the issues discussed in the above return then in a much richer way. A benchmark in the area that has kept generating surprising new angles is the Backward Induction algorithm (Aumann 1995, Bonanno 2001, van Benthem & Gheerbrant 2010). Many current logics of strategies can define how this works, but there are intriguing issues in interpreting what the mixture of action and preference in these logics achieves, and what becomes of the notion of a strategy in this setting. Proposals range from generating strategies as advice for best behavior to viewing strategies as beliefs about the behavior of other agents. All this gets even more complex when we go to techniques like Forward Induction that also take into account what history of play has gone on up until the present moment. For more on the state of the art in logical models of such reasoning styles, see van Benthem 2012B.

---

[7] This view of knowing a strategy under other circumstances seems an interesting counterpart to counterfactual views of knowledge in the philosophical literature that make knowledge a true belief that would also have been correct if the world would have been slightly different (cf. Nozick).

[8] Similar issues arise in analyzing what it means for someone to understand a formal *proof*, and useful intuitions might be drawn from our experience with mathematical practice.

Zooming in and zooming out This may be a good place for a brief digression clarifying what I am advocating with a logical point of view on strategies. My discussion of non-trivial game solution methods like Backward Induction may have suggested that logics of strategies must get ever more expressively powerful, making everything explicit, zooming in on the tiniest details. But that would be only one half of the story that I have in mind for the area of our Workshop. Logic may just as well do the opposite to get greater clarity, zooming out to a convenient abstraction level that hides details of a given type of strategy. In particular, it has been argued that practical reasoning needs coarse-grained modal top-level logics of *best action* as a subset of all available moves in a game (van Otterloo), and similar ideas occur in recent logics that merge ideas from game theory and deontic logic (Kooi & Tamminga, Roy). Indeed, several logical abstraction levels can make sense for one and the same reasoning practice, and the case of strategies is no exception. [9]

Architecture of diversity I conclude with two points that return to my concern at the start of this paper, the coherence of the area represented at our Workshop, and indeed the topic of this book. The first is one striking feature to any observer of our event, and a challenge to its organizers: the great diversity of paradigmatic scenarios that clamor for our attention, and of formal frameworks for dealing with these scenarios that compete for our allegiance. As to the latter, I am a firm believer in framework compatibility and convergence, but this methodological point is not something that I will argue here (cf. van Benthem, Gerbrandy, Hoshi & Pacuit 2009, van Benthem & Pacuit 2012 for a few samples). The less ideological issue is the architecture of the field of phenomena we are studying, which already showed in the diversity of terminology (strategies, protocols, tactics, etc.) mentioned earlier. The latter diversity is inevitable, as different domains may require different kinds of strategies. Unification of all these into one format, logical or otherwise, may not be possible, or even desirable. But even without aiming for unification, there is a challenge in achieving beneficial co-existence. Given that we use plans (pure or with information and preferences) in a wide array of differently structured tasks, what is the overall architecture that allows these plans to mesh and collaborate? It would be good to have a better understanding of ways of

[9] In terms of cognitive reality, zooming out and hiding procedural detail may mirror the cognitive phenomenon of *automation* turning explicit skills into unconscious routines in the brain.

interfacing and connecting strategies, harmonizing plans for short-term tasks (like in classical game theory) with those for long-term tasks (like in evolutionary game theory), or when on the same time-scale, making them work in parallel. There may be hidden complexities in this connecting up, just as with combining logics. And even more challenging than that, understanding the total architecture of strategies at these various levels may call for interfacing different mathematical paradigms, such as the dynamic logic of programs that we have highlighted earlier with the probabilistic dynamical systems of infinite games.

Giving the final word to reality I have raised a number of general theoretical issues that seem to run through the area of strategies as studied in this book. Many of them may seem common problems rather than shared solutions. But problems, too, can be a powerful bond in shaping a community. Let me add with a word of consolation all the same. Sometimes, when theoretical analysis seems to make things more, rather than less complex, it becomes time to think outside the box, in our case, going to the theorist's last resort of consulting the *facts*. I feel that the topic of our Workshop reflected an undeniably existing human practice. Conscious planning has been claimed to be the human evolutionary feature par excellence. Saving the last word for empirical reality, then, it would be good to bring in inspiration from cognitive studies of strategic behavior (Verbrugge 2009), since that is where our subject is anchored eventually. Now, making a significant connection here need not be simple. It may not be easy to make plan structure with its delicate compositional, generic, and counterfactual aspects visible and testable in actual psychological experiments. But that just means that, in addition to its current logical, computational, and philosophical dimensions, the topic of our Workshop also invites sophisticated empirical fact gathering,

References (to be expanded)

S. Abramsky, 2000, 'Concurrent Interaction Games', in J. Davies, A. W. Roscoe & J. Woodcock, eds., *Millennial Perspectives in Computer Science*, Palgrave, 1–12.

S. Abramsky & R. Jagadeesan, 1992, 'Games and Full Completeness for Multiplicative Linear Logic', *Journal of Symbolic Logic* 59:2, 543–574.

M. Birkegaard Andersen, Th. Bolander & M. Holm Jensen, 2012, *Conditional Epistemic Planning*. JELIA 2012. To appear in Lecture Notes in Artificial Intelligence, 2012.

R. Aumann, 1995, 'Backward Induction and Common Knowledge of Rationality', *Games and Economic Behavior* 8, 6–19.

R. Axelrod, 1984, *The Evolution of Cooperation*, Basic Books, New York.

J. van Benthem, 2012A, 'In Praise of Strategies', in J. van Eijck & R. Verbrugge, eds., *Games, Actions, and Social Software*, Lecture Notes in CS 7010, Springer, 96–116.

J. van Benthem, 2012B, *Logic in Games,* The MIT Press, Cambridge Mass.

J. van Benthem, J. Gerbrandy, T. Hoshi & E. Pacuit, 2009, 'Merging Frameworks for Interaction', *Journal of Philosophical Logic* Volume 38:5 (2009), 491–526.

J. van Benthem & A. Gheerbrant, 2010, 'Game Solution, Epistemic Dynamics, and Fixed-Point Logics', *Fundamenta Informaticae* 100, 19–41.

J. van Benthem & E. Pacuit, 2012, 'Choices, Actions, and Games', to appear in *Nuel Belnap Volume*, Trends in Logic, Springer.

G. Bonanno, 1993, 'The Logical Representation of Extensive Games', *International Journal of Game Theory* 22, 153–169.

G. Bonanno, 2001, 'Branching Time, Perfect Information Games, and Backward Induction', *Games and Economic Behavior* 36, 57–73.

R. Fagin, J. Halpern, Y. Moses & M. Vardi, 1995, *Reasoning about Knowledge*, The MIT Press, Cambridge Mass.

J. Hofbauer & K. Sigmund, 1998, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge.

M. Huth & M. Ryan, 2004, *Logic in Computer Science*, Cambridge UP, Cambridge.

R. Moore, 1985, 'A Logic of Knowledge and Action', SRI International, Menlo Park.

R. Nozick, 1981, *Philosophical Explanations*. Harvard University Press, Cambridge Mass.

M. Osborne & A. Rubinstein, 1994, *A Course in Game Theory*, The MIT Press, Cambridge (Mass.).

R. Ramanujam & S. Simon, 2008A, 'A Logical Structure for Strategies', *Proceedings LOFT 7*, Amsterdam University Press, *Texts in Logic and Games* 3, 183–208.

R. Ramanujam & S. Simon, 2008B, 'Dynamic Logic on Games with Structured Strategies', *Proceedings KR 11*, AAAI Press, 49–58.

Y. Venema, 2006, 'Algebras and Co-Algebras', in P. Blackburn, J. van Benthem & F. Wolter, eds., *Handbook of Modal Logic*, Elsevier, Amsterdam, 331–426.

R. Verbrugge, 2009, Logic and Social Cognition', *Journal of Philosophical Logic* Volume 38, 649–680.