Logic in Play

van Benthem, J.

Publication date
2017
Document Version
Author accepted manuscript
Published in
Logic, Methodology and Philosophy of Science

Link to publication

Citation for published version (APA):
van Benthem, J. (2017). Logic in Play. In H. Leitgeb, I. Niiniluoto, P. Seppälä, & E. Sober (Eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the Fifteenth International Congress (Helsinki) : models and modelling* (pp. 3-39). College Publications.

1 Logic in play

JOHAN VAN BENTHEM [*]

Abstract. Going beyond the traditional focus on consequence and inference, logic can be broadened to an exact theory of general information-driven agency drawing on many sources, without giving up on its well-established mathematical modus operandi. We show how this broader agenda involves the design of new kinds of dynamic logics for action, information, knowledge update, and belief change, and eventually, entangled with these, agents' preferences and goals. In particular, we explore several active interfaces of logic and games, where all these themes come together in natural concrete scenarios, ending up with advocating a move from game theory to a theory of play. Our presentation throughout takes the form of discussing typical examples, identifying major themes and suggesting new open problems concerning logic and agency. Finally, we explore what taking this agency-based perspective means for a variety of fields, including philosophy, linguistics, and game theory—and we conclude with some thoughts on what the field of logic is, or can become, in the perspective presented in this article.

Keywords: Dynamic-epistemic logic, games, agency, information.

1 Introduction: two perspectives on logic

What logic is about and what it is up to can be viewed in two different ways. One can see logic as describing the core structures of reality, with logical constants mirroring the structure of compound facts. In this perspective, logic is close to metaphysics, though some people soften this stance by letting logical consequence refer to infor-

[*] Amsterdam, Stanford & Tsinghua, http://staff.fnwi.uva.nl/j.vanbenthem

mational dependencies in the world, and logical constants to how we classify that information in ways that are useful to us. On either variant, logic is about structures out there in the world, structures that would be there even if no living being ever existed in any planet of this universe. This view of logic has a long and distinguished history going back to metaphysical or scientific inquiry. But there is also a second view where agents are of the essence, and it has an equally long pedigree in the history of the field. Logic is also typically embodied in human activities such as conversation or argumentation, and some historians believe that this is even how the scientific subject arose in the first place, out of reflection on this practice in philosophical, legal or political settings. Logical constants are then about structured moves that can take place in such social scenarios, while logical consequence has to do with forcing one's interlocutor to accept certain propositions. In other words, logic can be about the world out there, but just as well about the agents perceiving that world and acting in it.

The two perspectives are not at odds. Clearly, agents will only behave successfully in the long run if their modes of representation and reasoning fit the facts of the world. Also, the two views occur entangled in natural language, the culture medium for all we say and do. One instance is the pervasive 'product–process ambiguity' (van Benthem, 1996) between expressions for activities and their products in the world. Consider Carnap's famous book *Der Logische Aufbau der Welt* (Carnap, 1928). In German, "Aufbau" is ambiguous between 'structure' of the world and our 'construction' of it—and something similar is true for other natural languages. And these two viewpoints are clearly involved in what might be seen as an intriguing conceptual dance.

2 Logic and games: a natural combination

Despite endorsing this lofty balance, my purpose in this paper is to explore the activity or process perspective on logic, as it still has not received the full attention than it deserves (van Benthem, 2011). And going that way, actors come to the fore, that is the agents employing logic. So, my main topic might be called logic and agency.

In pursuing this line, I am going to first restrict attention to an area where many issues become concrete because we have vivid intuitions about them, namely, *games*. Games are a natural prism for the themes of this paper (van Benthem, 2014a). First of all, they are a natural practice where we hone our logical skills, and in particular, major logical activities such as argumentation have clear game-like features, such as choices of what to say, long-term strategies for dealing with opponents, and preferences as to the outcome of a debate. But also, games are a concrete model of intelligent social interaction, and they exhibit many structures that invite logical analysis. The interface of logic and games (and game theory) is rich and growing, with computer science as a 'Dritter im Bund', and we will show how. After that, I will explore more general agency-related themes and their connections with logic, and having done that, I conclude with some consequences of taking this agent-oriented perspective for a range of issues from philosophy to the sciences, and for logic itself.[1]

3 Evaluation games in logic

Let us make a simple connection first between a basic notion in logic and one in games, that has been proposed by a range of authors since the 1960s. Truth and falsity for formulas in models can be analyzed in terms of a two-player game, where we pull apart logical notions into different roles—a standard pattern in creating 'logical games'. Roles allow the mind to play against itself, testing things to the utmost.

Games, roles, and moves Let φ be a first-order formula, and $\boldsymbol{M}$ a model. Verifier claims that φ is true in $\boldsymbol{M}$, Falsifier claims that it is false. To be fully precise qua first-order semantics, we would also need an assignment sending individual variables to objects in the model, but in our exposition here, we will mostly downplay this finesse. Now the logical structure of the formula φ induces a scheduling for the game:

> At disjunctions, Verifier has to choose a disjunct for further play, while at conjunctions, it is Falsifier who has to make this choice. The game for a negation $\neg\varphi$ is the dual of the game for φ, all marks for turns, winning and losing are reversed between the two players. Verifier has to choose an object in $\boldsymbol{M}$ for an existential formula $\exists x\varphi$, while Falsifier chooses an object in $\boldsymbol{M}$ for a universal formula $\forall x\varphi$.

Each round drops a logical operator. When we reach an atom, a check takes place against the model $\boldsymbol{M}$, and the game ends: Verifier wins if the atom is true, and Falsifier wins if it is false.

Example A formula in a network.

Consider the network depicted in Figure 1, with five nodes, and a connected relation: that is, there is an arrow between every two points (though we will not draw the resulting reflexive arrows, for convenience). The first-order formula $\forall x\forall y\ (Rxy \vee \exists z\ (Rxz \wedge Rzy))$ says that one can get from any point in the graph to any other point by at most two directed arrows. This assertion is false in the given network, and the game will show how. For instance, Falsifier can pick the objects $x = 5$ and $y = 4$, and can then win against any counter-play by Verifier.[2]

There is something general going on here that connects logic and game theory.

Truth, falsity and winning strategies The following equivalence can be proved easily by induction on the structure of first-order formulas, and the definition of the game in a

[1] This paper is largely a programmatic survey, with an emphasis on ideas and suggestive examples. Hence, I will not include extensive references, which are given, for instance, in the books cited here.

[2] Note how games end in finitely many steps, since the formula in play gets smaller in each round.

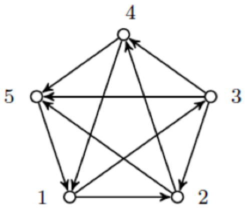

Figure 1.

model (the details of this proof again involve the use of variable assignments).

> *Fact* A first-order formula φ is true in a model $\boldsymbol{M}$ iff Verifier has a winning strategy in game(φ, $\boldsymbol{M}$).

This ties truth, a basic logical notion, to the game-theoretic notion of a strategy (for Verifier). Likewise falsity matches existence of a winning strategy for Falsifier.

Discussion Simple as it is, the preceding result suggests radical thoughts. The notion of a strategy is more fine-grained than truth, as there can be more than one winning strategy for a player. In the preceding example, another winning strategy for Falsifier is $x = 2$, $y = 3$. Thus, a game semantics is more fine-grained than mere truth values, a feature not exploited much so far, which fits intuitions about there being a natural hierarchy of less or more fine-grained denotations for sentences of a language.

Also, in the definition of the game, the clause for the quantifiers looks different from that for the connectives, challenging a standard analogy. A quantifier episode consists of two things: the choice of an object, and then, using that, playing the rest of the game. The real game operation here seems to be the sequential composition of a separate quantifier sub-game and the one for the remaining formula. This rearranges the standard geography of logical constants: the basic games are now atomic tests and object selection, while the general game constructions over these are choice, role switch, and composition. We refer to van Benthem, 2014a for the effects of this reappraisal. For instance, strikingly, the basic proof system underneath first-order logic from a game-theoretic point of view is then decidable.

Evaluation games exist for many logical languages. However, not all of these are as simple as the one we just showed. For instance, evaluation games for so-called fixed-point logics that can represent recursive definitions, say, of transitive closure or of wellfounded relations, allow for infinite histories of the game, since unfoldings of fixed-point variables in a formula under consideration may return to a larger formula of a shape encountered before. For such infinite games, the above lemma still holds, but the proof becomes much more delicate. This setting leads to deep connections with Automata Theory that we cannot go into here.

Excursion: model comparison There are many games for other logical purposes, such as testing satisfiability, or comparing models. We add an example of the latter to show another aspect of the intimate connection between games and logical syntax.

> Consider two models ***M***, ***N***. Player ***D*** (*Duplicator*) claims that ***M***, ***N*** are similar, while (*Spoiler*) maintains that they are different. Players set some finite number k of rounds for the game, the severity of the probe. In each round, ***S*** chooses a model, and picks an object d in its domain. ***D*** then chooses an object e in the other model, and the pair (d,e) is added to the current list of matched objects. After k rounds, the object matching is inspected. If it is a *partial isomorphism*, ***D*** wins; otherwise, ***S*** does.

A telling example compares the ordering of the integers and rationals: the latter a dense structure, the former discrete. Here is how this comes to light in the game:

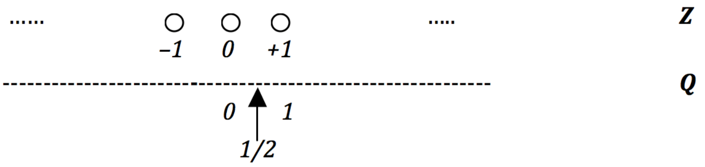

Figure 2.

By choosing objects well, ***D*** has a winning strategy here for the game over two rounds. But ***S*** can always win the game in three rounds, as suggested by our picture.

Here is a typical play:

Round 1	***S*** chooses 0 in ***Z***	***D*** chooses 0 in ***Q***
Round 2	***S*** chooses 1 in ***Z***	***D*** chooses 1 in ***Q***
Round 3	***S*** chooses $1/2$ in ***Q***	any response for ***D*** is loosing

In playing the games, winning strategies for ***S*** are tightly correlated with first-order formulas φ that bring out a difference between the models. It is easy to see how ***S***'s winning strategy in the preceding example matches step by step with the logical formula defining density:

$$\forall x \forall y\, (x < y \to \exists z\, (x < z \land z < y))$$

The analogy goes right down to how quantifier alternations mark model switches.[3]

Game operations and logical constants The structure of logical games may well involve further operations beyond the simple choices, switches and sequential composition that we have seen so far. In particular, there are also *parallel* constructions where different games are played at the same time. In fact, the preceding model comparison games are already a sort of parallel composition of games played in different models, with switches between these models initiated by one of the players. In general, on

[3] Model comparison games can also be continued infinitely, but we do not pursue this here.

the current view, the traditional set of logical constants can be naturally extended to mirror a broad spectrum of game operations.

4 Logic and game theory, basic encounters

One immediate effect of the above junction is that logical laws acquire game-theoretic import, and start connecting up with basic issues in game theory.

Excluded middle Consider the classical law of Excluded Middle $\varphi \vee \neg\varphi$. Its validity says that in every game of the form $\varphi \vee \neg\varphi$ over any model $\boldsymbol{M}$, Verifier has a winning strategy. But by our game rules, that strategy consists of a choice which game to play, and in which role. Unpacking this information, always, either Verifier or Falsifier has a wining strategy in the φ–game. Games having this very special property that one of the two players has a winning strategy are called *determined.* And we can see why first-order evaluation games have this property by referring to what may well be the earliest result in game theory.

Zermelo's Theorem The following result by Zermelo dates back to the 1910s.

> *Fact* Two-player zero-sum games with finite depth are determined.
>
> *Proof.* For each specific depth, and two players i, j, this is Excluded Middle unpacked into its game meaning. To see this, here is the case with two rounds: $\exists x \forall y WIN_{\mathrm{i}}xy \vee \neg\exists x \forall y WIN_{\mathrm{i}}xy$ is equivalent by pure logic to $\exists x \forall y WIN_{\mathrm{i}}xy \vee \forall x \exists y \neg WIN_{\mathrm{i}}xy$, which in its turn is equivalent, at least in zero-sum games, to a determinacy statement $\exists x \forall y WIN_{\mathrm{i}}xy \vee \forall x \exists y WIN_{\mathrm{j}}xy$.

However, there is also a generic proof across models solving the game by an algorithm that computes colors White for nodes where player *i* has a winning strategy and Black for nodes where the opponent *j* has a winning strategy. The algorithm colors end nodes according to who wins the game there. Next, working upwards in the game tree, it colors a node that is a turn for player *i* white if there is at least one white daughter node, and black if all daughter nodes were colored black. The coloring rule for turns of player *j* is the obvious dual.

Equilibrium and fixed-point logic First-order evaluation games satisfy the conditions of Zermelo's Theorem, and so, their determinacy is explained. For later reference though, note one can also look differently at the structure of the game solution process going on here. The coloring algorithm itself has a logical form that can be read as an inductive definition of the eventual winning-strategy predicates WIN_{i}, WIN_{j}. The stages of the algorithm correspond to steps of unfolding the recursive definition, until the first fixed-point is reached where the predicates no longer change. The shape of the driving formula here is as follows:

$$WIN_{\mathrm{i}} \leftrightarrow (end \wedge win_{\mathrm{i}}) \vee (turn_{\mathrm{i}} \wedge \langle move_{\mathrm{i}} \rangle WIN_{\mathrm{i}}) \vee (turn_{\mathrm{j}} \wedge [move_{\mathrm{j}}]\, WIN_{\mathrm{i}})$$

The right-hand formula has only positive syntactic occurrences of the variable for the strategic winning predicate WIN_i being defined here. Therefore, in standard fixed-point logics, we can prefix a greatest fixed-point operator to the right-hand formula to describe the eventual solution predicate.

This is not just a technical observation. Games are considered solved in game theory when we have an equilibrium between what players can achieve. Algorithms solving games often approach these equilibrium states in a stepwise manner. Thus, game-theoretic equilibrium is naturally connected with not just logic but also computer science, and the latter two fields meet in fixed-point logics of recursion and computation such as the modal μ–calculus, or $LFP(FO)$: the extension of first-order logic with operators for smallest and greatest fixed-points.

Logic and game equivalence But there is much more to the interplay of logical laws and games. Our second example again starts from a simple law of propositional logic, this time, Distribution of conjunction over disjunction. We display the two formulas involved in this law, and draw the shape of their game trees in a picture:

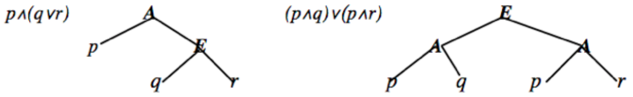

Figure 3.

Now we ask a new question, triggered by the logical equivalence.

Are these two games the same?

Distribution suggests that the two games are the same, and one can see that game-theoretically by focusing on the *powers* of players, i.e., those sets of outcomes that they can force the game to end in when they play one of their strategies against any possible counter-play by the opponent. It is easy to calculate that players have the same powers in both games.[4]

But there are natural alternative views when we look at players' actions and choices. The two games depicted are clearly different qua structure of moves and turns. For instance, there is no intermediate node in the game on the right that matches the situation in ***E***'s choice point on the left. To match this richer view of the 'how' rather than just the 'what' of control over outcomes, a more discriminating structural equivalence for games is needed. Good candidates for such a finer view are versions of the modal process equivalence of *bisimulation* that correlate available moves at each pair of matched nodes. Using a bisimulation, players can match their moves in two games

[4] In both games, ***A*** has the powers $\{p\}$, $\{q,r\}$, while ***E*** has the powers $\{p,q\}$, $\{p,r\}$. Here we use a condition of Monotonicity: powers of players are closed under taking supersets. Incidentally, powers in infinite games are sets of histories, but the idea remains the same.

step by step, thereby simulating the strategies themselves, not just the powers over final outcomes that they give rise to.

Further equivalences These are just two natural levels. Dropping monotonicity, van Benthem, Bezhanishvili, and Enqvist, 2016 identifies strategies with 'exact powers', where an outcome set shows what a player can force the game to end in, while the elements of that set show which choices are left to the *other player*. This notion of game equivalence differs from the earlier two: e.g., it does not validate Distribution. Thus, it induces an interesting weaker propositional logic yet to be explored. It also supports new game languages of the sort explored in the next section.

Thus, in this simple setting, we encounter the intriguing possibility that logics with different validities may match up with different structure levels for looking at games. In the following section, we will connect logic and games in one more manner.

5 Invariances, languages, and zoom levels

.

What we see here is an instance of a general phenomenon in mathematics and other fields. A subject can be studied at various levels of detail, and these levels are specified in terms of *invariance* under suitably chosen transformations, as proposed by Helmholtz and Klein in the 19th century, and in 20th century logic, by Tarski. There need not be a unique best choice here: Euclidean geometry is not 'better' than topology, it all depends on one's purpose. The same is true for games, and indeed, game theory has various natural structure levels, from extensive game trees to strategic matrix forms closer to the above power level.

Of special interest to logicians is that setting a level of detail corresponds to introducing a *language* that defines just the invariant properties characteristic for the chosen level.[5] The finer the invariance relation, the richer the matching language.

A typical illustration are the above two views of games. The richer action–choice level suggests a standard *modal language* over game trees viewed as relational models, where we have many results showing that (with some technical caveats) bisimilar models share the same modal theory. By contrast, focusing on powers suggests a coarser (but interesting) modal 'neighborhood language' for invariant properties, with 'forcing modalities' describing properties of players' powers. So, logical analysis of games is not embodied in one unique formal system: there is a hierarchy of logical languages and their logical laws matching a hierarchy of natural levels for representing games, or social interaction generally.[6]

[5] Helmholtz' own original motivation was finding an underpinning for the language of geometry.

[6] Our earlier example of Distribution suggested that logical equivalence is based on the power view. This

One can also restate these points in terms of 'zoom'. Many people believe that logic is organized pedantry. We take a given reasoning practice, and then supply more and more details until all arguments are fully spelled out, say, the way mathematical proofs might be spelled out to machine-readable first-order formulas. This is the *zooming-in* direction of supplying ever more detail, allowing us to see new phenomena at a microscopic level, such as small proof steps that can be automated. On this view, logicians are like moles digging in the soil below the observable cognitive behavior. But there is also a *zooming-out* direction where logical analysis does exactly the opposite: one looks at a reasoning practice, but only considers some global features in a rough formalism—the way, say, modal logic can yield a bare theory of the topological interior operation. This time, we see new things precisely because we ignore details, and soar, free as birds, far above the given reasoning practice. Both directions of zoom occur in logic, and there are interesting conceptual and technical questions concerning a systematic back-and-forth among various levels of analysis.[7]

In the light of Sections 4 and 5, developing a rich multi-level view of games and corresponding logics seems a well-worth enterprise.[8]

6 Entanglement in two directions

Our discussion so far has shown that logic and games form a natural combination. But, stepping back, we have really presented a mixture of two different directions.

Logic of games In one direction, often called logic *of* games, we use techniques from logic to analyze structures in games. In this contact, logic is applied as it stands, and the resulting systems are sometimes called 'game logics'. This is the direction that can be found in logical theories of multi-agent systems (Shoham & Leyton-Brown, 2009), or in current work on the logical foundations of game theory, as in 'epistemic game theory' that analyzes players' reasoning and strategic equilibria by logical means (Branderburger, 2014; Perea, 2012).

Logic as games But there is also a converse strand through many of the examples presented so far, where games themselves are used to analyze basic notions of logic, or in a current phrase, we pursue logic *as* games. This second direction of thought can run from weak claims, where 'logic games' are just used as convenient didactical tools, to strong methodological programs where games are viewed as embodying the essential meaning of logical systems.

may pose an initial perplexity, but we will return to this issue in Section 18 below.

[7] See van Benthem, 2016 for definitions of the phenomenon of tracking between levels, and some general results on when it works and when it does not, including connections with Category Theory.

[8] We will revisit this issue later in our discussion of 'Theory of Play', since our intuitions about game equivalence may have a hidden parameter: the type of agents that are playing the game.

Cycles The two directions are not at odds, although one should beware of confusion. For instance, in our game-theoretic analysis of first-order formulas, we can view these formulas as statements about some given model, but also as algebraic terms defining a kind of game playable on any model. These games have properties which can themselves be stated in some further logical (meta-)language, which could itself have evaluation games, and so on.

Thus, there is a productive cycle here. For instance, one can start from a class G of games, introduce a logical description language $L(G)$, and then consider games for evaluating the formulas of that logic, or for comparing its models. These activities in $G(L(G))$-mode are not always disjoint: a model comparison game for a logic of games may be close to a notion of structural equivalence for games in G. But $L(G(L))$-mode makes sense just as well. For instance, we defined evaluation games for first-order logic, but in studying those games from a solution perspective, we found patterns that suggests a natural fixed-point logic for those games, whose expressive power is known to exceed that of first-order logic itself.

The two directions are not as disjoint as they may seem. The monograph van Benthem, 2014a devotes two whole parts to hybrid systems merging motivations from logic games and game logics.[9]

Computation Our main point here is that, while logic and games form a natural combination, they do so in different entangled ways. Nevertheless, this two-component picture may also be misleading as it underplays what may well be an essential third ingredient: the role of *computation*.

Much of the basic work on game logics and logic games today is happening in the foundations of computation (Grädel, Thomas, and Wilke, 2002, Abramsky, 2008), where infinite games, Automata Theory, and Co-Algebra enter the fray, and where the emphasis in studying computation is shifting from extensional input-output views to the production of behavior by interactive systems. This is congenial to the later turn to be made in this paper toward general agency, since today's interactive computation is really a form of social agency (van Benthem, 2015), blurring the border line between games and computation just as much as that with logic.

Perhaps the best eventual picture is a triangle of *Logic, Games, and Computation*.

7 From logic and games to intelligent agency

In the rest of this paper, logic of games played by agents will be the central theme.

However, let us emphasize that logic as games is a rich area, too. In particular, our few examples should not be taken to convey the whole flavor. Evaluation games are

[9] There are even challenges to the whole scheme: for instance, evolutionary games seem hard to fit in.

just one way of casting logical notions as games. Argumentation, too, is a game, and winning strategies in suitably defined formal argumentation or dialogue games can be identified with proofs, in a tradition going back to Lorenzen. Argumentation or dialogue is also a powerful model for interactive computation, and the resulting work in the foundations of computation is producing deep results about games with a general thrust. For instance, the category-theoretic treatment of 'game semantics' in Abramsky, 2008 brings to light new logical constants, such as sequential versus parallel readings of disjunction that obey different laws. All this links up with resource logics such as linear logic, Proof Theory, Type Theory, and many other fields. Again we see how computation is a natural partner here.[10]

Having found natural interactions between logic and the study of games, we will now broaden our scope. Games are played by *agents*, and these agents are involved in a wide variety of activities. These range from deliberation before the game starts to processing information, or even dealing with unexpected surprises, during the game, all the way to post-game analysis, and perhaps rationalization of one's behavior.

In the coming sections we discuss a wide range of basic abilities of agents, relevant to playing games but also much more broadly, that can all be studied in logic.[11]

8 Dealing with many information sources

Inference and questions Inference is an important source of information, but there are others on a par with it. Suppose you are in a restaurant.

> Three people have ordered drinks, one each: water, beer, and wine.
> A new waiter comes carrying three glasses. What happens?

There are six ways the glasses can be distributed over the three customers, and here is a scene that plays out every day in many places. The new waiter needs to reduce the 6 options to 1, and solves his information problem as follows. He asks who has the water, puts that glass, then asks who has the beer, puts that, and then, without asking, puts the wine. What we see here is questions and answers reducing uncertainty from 6 options to 2 and then to 1, with the last step just being an inference (either explicit, or implicit in the act of placing the glass of wine).

There is a unity in this scenario: questions and inferences go together.

[10] For a broad panorama of the two directions at the interface between logic and games, with a landscape of game logics at different zoom levels, a presentation of different ways in which laws of logic can embody game-theoretic principles, and a survey of game-like systems that show features of both directions distinguished here, we refer to van Benthem, 2014a.

[11] As always, our exposition is a survey, for further details and results we refer to van Benthem, 2011.

Three sources of information There are further natural informational acts. This is already clear in the natural sciences: experiments and observations count for as much as mathematical deductions. An elegant compact statement from the world of practical common sense can be found in ancient Chinese Moist texts (400 BC):

知 闻 说 亲 zhi wen shuo qin

This elegant sparse phrase says that knowledge (zhi) comes from: hearing from others (wen), proof (shuo), or experience (qin). The Moist illustration is of someone seeing an object inside a dark room, wondering about its color. He sees a white object outside of the room, and someone tells him that the object inside the room has the same color as the one inside. He then infers that the object inside is white. What we see is a cooperation between observation, communication, and inference.[12]

Our basic informational abilities have at least this threefold range, and logical theory should deal with modeling all of this. As we will see soon, it can. But for now, let us continue with one more striking aspect of human reasoning abilities.

9 Social interactive reasoning

To some people, the high point of rationality is embodied in Rodin's 'Thinker': eyes closed, all on one's own. But in reality, intelligence seldom comes alone. In nature, many-body problems are the key to how the universe works, rather than the 'natural place' of individual objects. Likewise, our intelligent abilities usually unfold interactively in contacts with others. This so-called 'theory of mind' is seen as a crucial human ability in cognitive science, and it reaches far beyond the world of practical common sense: even the purest sciences themselves are a social enterprise.

Multi-agent knowledge 'Many-mind problems' are already key to asking the questions of our earlier base example. In addition to ground-level facts, we also communicate what we know or do not know about others. When I ask you a question, I normally convey to you that I do not know the answer, and also, that I think that you may know the answer: the latter is iterated two-agent knowledge.[13] Answering a question conveys a fact but, if done in public, it also makes sure that both participants know that they now know the answer, and this knowledge even goes to higher depths of entanglement, all the way to 'common knowledge' in the group.[14] It is this sort of iterated and entangled knowledge and ignorance of various sorts that keeps communication flowing, but also holds it in place and makes it successful.

[12] Frege famously gave up on natural language in favor of his "Begriffsschrift" because of the 'prolixity' of the former. But by 'natural language' he meant German: what if he had known Chinese?

[13] Exceptions are rhetorical questions, say, by teachers. But we are also attuned to when these occur.

[14] Of course, there is more to questions than just conveying information: tyically, questions also raise *issues*, and they set or modify a current agenda of conversation or inquiry.

There is logical structure underneath all this. We can reason about knowledge of agents about ground facts and each other in systems of *epistemic logic*, even group notions like common knowledge then turn out to have a precise logical behavior.

Just for illustration, we state an important valid equivalence relating knowledge of individuals and that of groups. Here we write $E_G\varphi$ for 'everyone in the group G knows that φ' and $C_G\varphi$ for 'φ is common knowledge in the group G':

$$C_G\varphi \leftrightarrow (\varphi \wedge E_G C_G\varphi)$$

This is not exactly a game-theoretic formula of our earlier fixed-point type, but it has a similar logical form, and it expressess an informational equilibrium in a group.

This example also illustrates a much more general point that is often under-appreciated. Action and information can exhibit very similar patterns, at least in the light of logic.

Dynamics with cards But here is another fundamental aspect to be recognized in the above. The whole point of communication as a social process is that what agents know keeps changing all the time as informational events occur. Again, what happens in the course of a game provides many concrete appealing examples of this flux. Consider the following baby card game:

> Three players, John, Mary, Paul get one card each: John gets **red**, Mary **white**, Paul **blue**. Now Mary asks John: "Do you have the blue card?" Who knows what now? John answers: "No". Who knows what now?

Audiences differ on answers to this little puzzle, but many people manage to figure out that after the question, John (but not Paul) has learnt who has which card, while after the answer, both John and Mary (but not Paul) know the cards. However, the fact that John and Mary know is common knowledge, so even Paul has learnt a lot from the information exchange in this scenario.

10 Information change, update, and dynamic logic

As we shall see, logic has the resources for describing scenarios like this: its scope extends beyond inference to questions—and to observation, and other basic acts.

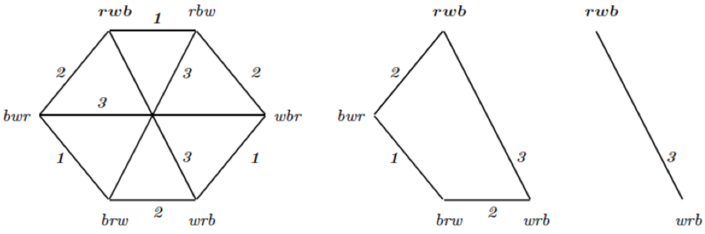

Figure 4.

Art of modeling We can represent the initial situation as a simple epistemic model with 6 states for the possible deals of the cards, and uncertainty lines indicating what agents cannot tell apart. For instance, in the actual state ***RWB***, Mary knows she has the white card, but cannot tell whether she is in ***RWB*** or in ***BWR***—and so on.

Now, crucially, further informational acts or events will change this initial model, following a widespread intuition in the literature that reflects common sense: new information decreases the current range of possibilities. For instance, learning that Mary does not have the blue card removes 2 options (***RBW***, ***WBR***) to yield the second picture shown in the sequence, where we can see graphically that in ***RWB***, 1 knows the cards: no uncertainty line departs there for him. The picture also shows many more facts in a direct visual manner: e.g., Paul knows that John knows the cards. The answer removes two more worlds (***BWR***, ***BRW***), resulting in the final stage depicted where the only remaining uncertainty line is for Paul. [15]

More generally, the informational action here is a public event $!\varphi$ telling everyone that φ is the case. (This is often called a 'public announcement' or 'public observation' of φ.) Its effect is the following change in the current group information state:

> it takes a current epistemic model ($\boldsymbol{M}$, s)—with an actual situation s —to
> a new model $M|\varphi, s$, where only those points remain that satisfy φ in $\boldsymbol{M}$.

Public announcement is about the simplest informational event imaginable. Understanding its logical behavior is the key to understanding a wide range of more sophisticated informational model updates.

Dynamic logics of information Informational acts $!\varphi$ satisfy precise logical laws, in suitably chosen languages allowing us to state what agents learn. The key principles of such systems describe the basic 'recursion equations' for the information flow triggered by acts or events, stating what happens to one's existing knowledge through an informational event. We display one such law:

$$[!\varphi]K_{\mathrm{i}}\psi \leftrightarrow (\varphi \rightarrow K_{\mathrm{i}}(\varphi \rightarrow [!\varphi]\psi))$$

This says that agent *i* will (would) know that ψ after receiving the truthful information that φ is the case if and only if, assuming φ holds, the agent had conditional knowledge of the implication that ψ will be the case after the φ–update.[16]

The reader may be used to laws of logic saying what agents know automatically when they know certain other things. That is an extreme static case, with information com-

[15] The semantic model drawn here matches sketches that many people make of the scenario, being a graphical representation of crucial information. (Incidentally, making good sketches from scenarios stated in natural language is an art going far beyond the routine translations into 'logical form' that we drill our students in.) But while we have used update only as the engine of information flow, actual behavior in problem solving also includes inference steps. Human reasoning seems a hybrid of many informational facilities, crossing boundaries that we would normally keep separate as theorists.

[16] The latter proviso with truth after the update, not simpliciter, is truly needed in a dynamic setting. We cannot just say $K_{\mathrm{i}}(\varphi \rightarrow \psi)$, because updates may change truth values of epistemic statements ψ.

ing 'for free'. Dynamic laws like the preceding generalize this to the wider informational setting highlighted here, telling us what agents know after they have taken the trouble to explore or communicate.

Multi-culturality Also noteworthy is the cooperation of several disciplines embodied in this one formula, the same way common artifacts of modern life, say, your reading glasses, are often little crystallized pieces of a long history of cross-cultural collaboration. The idea of having logics for knowledge ($K_i\psi$) comes from philosophy, representing speech acts $!\varphi$ explicitly can be seen as the essence of linguistics, and the methodology of describing change after events with modalities comes from dynamic logic of programs in computer science. This reflects the fact that for the program of this paper, traditional border lines between philosophical logic, computational logic, or mathematical logic make no sense. We need all the insight we can get if we are to understand the full realm of the logical.

11 Proven methods, broader scope

The preceding observation suggests a continuity that may be worth stressing. Extending the scope of logic as advocated here is not a break with the past. The standard methods and standards of rigor of logic still apply, and in fact, they are needed to get a grip on what is going on. Information dynamics has laws extending our usual repertoire, but its theory is still logic. In Section 21 of the paper we will return to the issue of how to view all this.[17]

12 From information processing to agency

So far, the agents that we describe are mere information-recording devices. But human agency is more than information processing. Behavior is driven by goals and preferences. Now to some people this means entering the realm of taste and arbitrariness, but in fact, there is logical structure even here when we analyze styles of reasoning. Tastes may be arbitrary, but the world of fashion can have very rigid laws.

Here is our running example of the complexities arising in the interplay of information and preference, even in extremely simple game scenarios:

Reasoning with preference In the following game, player ***A*** can go left, and get 1 Euro, while ***E*** gets nothing, or ***A*** can go right, giving the turn to player ***E***. Then ***E*** can

[17] As the reader will see, our main line does not involve changing the standard laws of classical logic: dynamics does not call for alternative logics, but for extended vocabularies. While this may provide fewer thrills than the vertiginous joy rides into the deviant, vague or inconsistent that some philosophers prefer, we believe that alternative logics should not be multiplied beyond necessity.

go left, getting 100 Euro while ***A*** gets nothing, or right, in which case both players get 99 Euro. What might happen in this scenario?

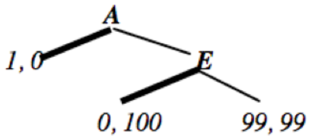

Figure 5.

One way of thinking is this. At her choice point, ***E*** faces a standard decision problem, and if she is rational, she will go left since she prefers that outcome over the one to the right. But ***A*** can see this coming, and realizes he will get 0 in that case, whereas going left will give him 1. So, ***A*** goes left, and it does not matter what ***E*** does.

This style of reasoning is called *Backward Induction*, and it describes a style of play for rational competing agents. Backward Induction has become a benchmark for logics of games, and we will return to its analysis below.

However, there are alternatives. Cooperative players might reach the endpoint (99, 99) by other plausible kinds of reasoning. There are many footholds for this. For instance, one weakness of the above argument is that it ignores the fact that ***E*** does not make her choice ex nihilo: she can also take into account the history of the game that led up to it, in terms of what she believes about the type of player that ***A*** is, or even if she has no such belief, she may just feel that 'she owes him'. In longer games than the one displayed, this could certainly matter.

In general, in most games we do not know which type of player we are up against—and the common distinction between 'competitive' versus 'cooperative' games does not help. For instance, academic life is a subtle (and sometimes not so subtle) mixture of both. So, we need an abstract stance that accommodates variety of behavior.

Philosophical plus computational logic Therefore, it makes sense to design logics that can account for any reasoning of the above kind about social scenarios, mixing information, action, belief, and preference. And when we analyze the ingredients for that reasoning, they read like a compact agenda for all of philosophical logic, involving knowledge, preference, belief (after all, ***A*** will never find out what ***E*** would do, so he cannot know it, but only have a belief about her decision), counterfactual conditionals (what would ***E*** have done, had ***A*** moved right), but also notions from computational logic in single actions or complex strategies, and fixed-points corresponding to various game-theoretic equilibria, such as the ones computed by Backward Induction, or by cooperative scenarios.

Again, we see our point of merging disciplinary agendas. Logic of games is an area where many different strands in earlier literature meet in concrete scenarios.

13 Benchmark for game logics: Backward induction

For a concrete illustration, we will look at the logical form of Backward Induction. Whether we endorse this style of reasoning or not, what is the logical form underlying its particular take on rationality? We will use this a showcase of our general approach in Sections 4 and 5: different zoom levels make sense for different purposes.

Backward Induction algorithm To define the algorithm, we first need an auxiliary notion. We say that a node x strictly dominates a sibling node y (siblings are immediate successors of some shared parent node z) for player i if all further outcomes of the game reachable after x are preferred by i to all outcomes of the game reachable after y. Now we can define a relational version of the BI algorithm as follows (the earlier numerical values were just added for drama): one keeps removing transitions from parent nodes to strictly dominated children. In general, this process must remove transitions, when we look at points whose only children are end nodes. Moreover, iteratively, the algorithm will move upward in the tree (it is really a refined version of the earlier Zermelo coloring algorithm) until we reach the root. Since the 'available move' relation BI can only get smaller in the process, it must stop by some stage, and this fixed-point is the solution produced by the algorithm.

> *Theorem*
> *BI* is the largest sub-relation of the *move* relation in finite game tree satisfying (a) the relation has a successor at each intermediate node, (b) *CF*:

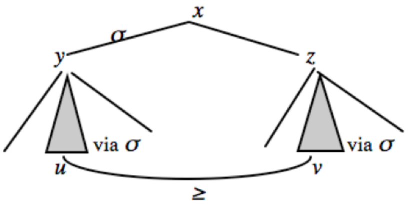

Figure 6.

> *Confluence (CF)*
> $\&_{\mathrm{i}}\ \forall x \forall y\, ((Turn_{\mathrm{i}}(x) \wedge x\sigma y) \rightarrow (x\, move\, y \wedge \forall z\, (x\, move\, z \rightarrow \exists u \exists v\, (end(u) \wedge end(v) \wedge y\sigma^{*}u \wedge z\sigma^{*}v \wedge v \leq_{\mathrm{i}} u))))$

The logical form of rationality In a typical logician's modus operandi, we can now look at a (first-order) syntactic recursive description of the algorithm. It is easy to see then that the relation R that is produced only occurs *positively* in the above formula with $\forall\forall\exists\exists$ syntax. Thus we have an observation which again reflects the close connection between game-theoretic equilibria and fixed-point logics.

> *Fact* The Backward Induction strategy, viewed as a sub-relation of the total *move* relation, is definable in first-order fixed-point logic $LFP(FO)$.

One can view the fixed-point formula found here as bringing to light the 'logical form' of rationality as conceived of by Backward Induction. Other game solution methods may induce other logical forms.

Domination and belief But there are further interesting features worth pointing out, on the following intuitive interpretation. What is 'dominated' can change at each stage since the total set of available moves gets smaller. This set of available moves, and hence available histories of the game, may be said to represent the players' *current beliefs* about how the game might still go. Avoiding dominated moves at the current stage, as prescribed by the algorithm, is then a form of optimal decision making *given one's beliefs*, or in game-theoretic terms: players are assumed to be 'rational-in-beliefs', regardless of whether they are doing what is objectively best for them.

This belief interpretation can be made into a dynamic reanalysis of Backward Induction as an iterated belief revision procedure for players engaging in pre-game deliberation. We will look at this mechanism later on, since it has uses in many places.

As one more pointer to later themes, note that the Backward Induction solution is *uniformly definable* for both players, who are assumed to exhibit the same kind of reasoning and global beliefs about how other players operate – even though their base-level preferences and the available moves at their turns may differ completely. Once we drop this assumption of uniformity, however, we have to start thinking about games in which players can be of different types, raising the issue of how much diversity one can tolerate before social interaction, or at least, reasoning underlying social interaction, breaks down. This theme will return in our section on 'Theory of Play'.

Zoom levels Finally, recall our earlier theme of different zoom levels for logical analysis. Our definition of the Backward Induction strategy in fixed-point logic is extremely fine-grained. It is hard to imagine that reasoning in such detail informs all our behavior in daily interactions. Can we zoom out to higher description levels?

One way of zooming out uses a modal language for 'best actions' and preferences only, with Backward Induction kept under the hood. How will such surface level reasoning go? The earlier rationality principle is now expressed by this modal axiom, where σ denotes optimal moves (best actions) recommended by the algorithm:

$$(\mathit{turn}_{\mathrm{i}} \wedge \langle\sigma^*\rangle(\textbf{\textit{end}} \wedge \varphi)) \rightarrow [\mathit{move}_{\mathrm{i}}]\langle\sigma^*\rangle(\textbf{\textit{end}} \wedge\langle \mathit{pref}_{\mathrm{i}}\rangle\varphi)$$

However, the complete modal logic of best action in this sense is not known, and it is not even clear that it is completely axiomatizable.[18] There are interesting analogies

[18] The technical reason is an insight from computational logic. The 'confluence patterns' in game trees induced by our BI analysis create a regular grid-like structure that may support embedding of so-called 'tiling problems' of high complexity in the logic. If this is so, then, though rational behavior itself may be simple and predictable, the logical theory of rationality might be highly complex.

between this modal level of analyzing best action and that of *deontic logic*, viewed as a high-zoom specification language for regulating and evaluating behavior.

The natural language of decision There may be other zoom levels here, and logic need not always have the best description language for interactive social behavior lying on the shelf. Indeed, natural language itself seems to do a good job in this respect. Our rich vocabulary concerning best or better actions, our hopes and fears, what we ought or are permitted to do, and other evaluative expressions form a rich and subtle medium for describing and influencing behavior. So far, logics of agency have only scratched the surface of this.

14 Players do much more

Three phases What we have looked at so far is only a tiny slice of the intelligent activities that players exhibit. Let us recall an earlier distinction. There are at least three different phases where logical reasoning plays a role: *before*, *during*, and *after* a game. The above Backward Induction procedure, and many 'game solution' procedures seem to belong to a pre-game *deliberation* phase. But also very important is the post-game phase of analysis and usually, *rationalization*, where we fix what is the 'lesson learnt' for future play. And perhaps most excitingly, many things happen during a game. Moves are played, obviously, but also, information is received and gets processed, as we saw in our cards scenario.

Adding the past One immediate instance of in-play reasoning occurs with Backward Induction itself. What if, in the course of a game, we see that our opponent deviates from the expectations generated by the BI algorithm? This time, in general, we have two inputs: what we expected beforehand, and what has actually happened.[19]

This opens up a wide space of options. We might consider the deviation an error. We might take it as a signal that the other player wants to cooperate with us, at least to some extent. We might also take a repeated simple pattern of deviations as information of a very different kind. Perhaps we are witnessing a drastic case of non-uniformity, and we are playing against an automaton that always plays one sort of move.

Belief revision It will be clear that the usual mathematical notion of a game tree, as a record of all possible runs of a game, does not suffice to model all these processes going on in play. Accordingly, various additional aspects have started appearing in studies of games, such as explicit modeling of 'player types' in epistemic game theory, or of various kinds of automata playing games in computational logic. We will return to this shift later, but for now, let us just note a more base-level common denominator behind the preceding scenarios. What surprising events in a game will do in general

[19] By contrast, our running example of the Backward Induction algorithm only looked at the future at any node, and then, we might just as well throw away the past play leading up to that node.

is force players to engage in on-line *belief revision*, leading to new expectations about the future course of the game.

This broader area is the realm of game-theoretic solution methods like ‘Forward Induction’.[20] However, in this paper, we will look at things in a more general logical perspective, tying up with one more general aspect of agency In the coming two sections, we discuss two additional aspects of agency that we did not high-light before.

15 Dynamic logic of belief revision

From knowledge to belief Our actions are driven by belief as much as knowledge. And beliefs are not whimsical attitudes, or the soddy paper money that mimics the gold standard of knowledge.[21] Beliefs are crucial triggers for most of our actions, and their formation and maintenance may even have a more creative aspect than the knowledge in one’s savings account.

Beliefs are generated by a much wider spectrum of informational events than the indubitable public announcements that we have considered so far. In particular, they can be triggered by signals carrying what might be called ‘soft information’ as much as by the hard information that we modeled earlier in our discussion of knowledge update in card games. More generally, changing beliefs is not a vice but an epistemic and cognitive virtue: our cognitive abilities often show at their best when we are confronted with a surprise, and need to re-adjust what we thought before.

Logic of belief revision Belief revision is crucial to the realm of agency that we are charting, and techniques similar to our earlier ones apply. First, we need a simple static base model for beliefs, and one common such structure are plausibility models $\boldsymbol{M}= (W, \sim, \leq, V)$, where the agents’ ranges in an epistemic model as defined earlier are now ordered by a binary relation of relative plausibility.[22] These models are a qualitative pilot setting on which to develop the main themes to follow—though they can be enriched with further structure (evidence, or probability) when needed.

Belief in the truth of a proposition φ (written $B\varphi$) can now be defined as truth of φ in all the most plausible worlds. But we need more expressive power. Given that information about the relevant set of worlds may change, we also need a notion of *conditional belief* $B^{\psi}\chi$ saying that χ is true in all most plausible worlds within the

[20] Such alternative solution methods may no longer work on simple annotated game trees. For a systematic hierarchy of logical models for games with increasing complexity, cf. van Benthem, 2014a. Richer models are needed as we relax what players know about the game and each other’s strategies.

[21] Conceptually, paper money may have been the more innovative historical invention.

[22] We drop agent indices in what follows, but this is merely for notational convenience. There is no barrier toward dealing with multi-agent scenarios.

set of ψ–worlds.[23] Plausibility models also support other epistemic attitudes, such as 'safe belief' and 'strong belief', but we will not pursue this theme here.

Hard information The static logic of belief defined in this way is much like conditional logic in the semantic tradition. Of interest to us here is that our earlier theme of information flow and change forms a natural continuation of such systems from philosophical logic. First, consider the earlier events $!\varphi$ that produce the hard information that φ is the case. These sinple events can already drive quite interesting scenarios, witness the 'misleading with the truth' cases discussed in van Benthem, 2011, whose details we forego here.

> *Fact* The logic of changes in absolute and conditional beliefs under hard information is completely axiomatizable using suitable recursion axioms.

We merely display the two basic recursion laws for new beliefs:

> *Fact* The following equivalences are valid for hard belief revision:
>
> $[!\varphi]B\psi \leftrightarrow (\varphi \rightarrow B^{\varphi}[!\varphi]\psi)$
>
> $[!\varphi]B^{\psi}\chi \leftrightarrow (\varphi \rightarrow B^{\varphi\wedge[!\varphi]\psi}[!\varphi]\chi)$

Soft information Belief is not just one more attitude that changes under hard information. There can now also be events that do not eliminate worlds (every existing option remains available), but modify the plausibility pattern. There are many actions of this kind in the literature, but it will suffice to mention one characteristic example:

> A radical upgrade $\Uparrow\varphi$ puts all φ–worlds on top of all $\neg\varphi$–worlds in the ordering, and within these two zones, it keeps the old ordering.

Radical upgrade is a strong move in favor of φ, of a sort that has been studied in belief revision theory (Gärdenfors & Rott, 1995) and formal learning theory (Baltag, Smets, & Gierasimczuk, 2011).[24]

Again the complete dynamic logic of such events, viewed as denoting matching model changes, can be described by introducing appropriate modalities.

> *Fact* The logic of changes in absolute and conditional beliefs under soft information is completely axiomatizable using suitable recursion axioms.

This time, we display just one, formidable-looking, recursion law, where 'E' stands for the existential epistemic modality 'somewhere in the current epistemic range':

$$[\Uparrow\varphi]B^{\psi}\chi \leftrightarrow (E(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$
$$\vee\ (\neg E(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

[23] More complex truth clauses are needed for infinite models. Note also that, unlike with the earlier conditional knowledge, conditional belief is not definable in terms of absolute belief.

[24] Alternatively, a radical upgrade can be seen as a strong deontic command to see to it that φ.

For an explanation of this law, and a method for its systematic derivation from the definitions of radical upgrade and conditional belief, see van Benthem, 2011.

Laws of learning The principle displayed just now shows that logical laws can describe the formation of new beliefs, and even of new conditional beliefs, as new information comes in. An alternative interpretation is as laws of *learning*, since much learning consists in modifying beliefs so as to improve their fit with the truth, or at least with reliable new information. However, there is nothing peculiar about radical upgrade that enables us to do this. Plausibility changes can come in a great variety of formats or ‘learning policies’. And there exist several general methods for dealing with the induced dynamic logics of belief change. The survey van Benthem and Smets, 2015 references many classical contributions to the literature.

Finally, while the above law may look much more complex than standard axioms for epistemic or doxastic logic, this is to be expected. Belief is a subtle notion—and the greater complexity may also be seen as greater richness of content.

We have seen now, at least as a sketch, how logical techniques can deal with describing beliefs and belief changes just as for knowledge. Since beliefs are fallible, and can be wrong, this shows that logic is not tied exclusively to truth and knowledge, it also makes sense as a guide when we live in a world of error and confusion. We will return to this theme of *correction* rather than (just) *correctness* in a later section.

Coda Our themes here do not belong exclusively to logic. Learning and improving theories have long been major themes in the philosophy of science, and the same broad origins can be seen with counterfactuals, or belief revision theories. But then, at an Auld Alliance Congress like the DLMPS, who cares about exclusive labels like ‘logician’ or ‘philosopher of science’?

16 Long term phenomena and limit behavior

A second prominent feature of games with logical import is their temporal horizon. *Strategies* are usually not single moves, but methods for achieving some effect only after many steps. Strategies in infinite games need not even work toward any finite apotheosis, but serve to produce particular kinds of never-ending *histories*. And in the temporal long run over histories, phenomena may emerge that are sui generis.

Limit phenomena A temporal perspective enriches our earlier dynamics. Single informative events, whether with hard or soft information, form longer histories. These may contain surprising emergent structure of their own, such as success or failure in converging to a fixed-point when running an update or deliberation procedure.

Consider the famous ‘Muddy Children’ puzzle (cf. van Benthem, 2011 and the references therein) where repeated announcement of the children’s ignorance of their status (as long as this ignorance holds) results eventually in a flip-flop: some children do know their status after the last announcement of ignorance. Here, in addition to immediate effects of the individual statements, we see ‘self-refuting’ limit behavior of iterated announcements: at the final stage, the statement becomes false in the actual world. But this is just a start. Limit behavior of assertions can also be ‘self-fulfilling’. Then, at the first fixed-point of iterated announcements, the statement announced is true everywhere in the remaining model, making it common knowledge among the agents involved. Here is a game scenario where this happens.

Example Backward Induction, hard scenario.

Consider our earlier analysis of the Backward Induction algorithm. Let ***rat*** be the statement — which can be true or false at any given node — that no player played a strictly dominated move when coming to this node. Announcing this information in our earlier public announcement style will leave certain nodes, but may remove others. Thus, afterwards, new nodes may satisfy, or fail to satisfy, ***rat***—and hence iterating this assertion makes sense.

Consider our earlier running example. The following figure depicts what happens in a hard scenario of events !***rat*** removing nodes from the tree that are strictly dominated by siblings as long as this can be done:

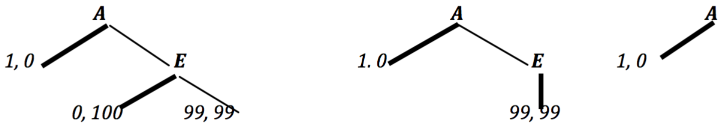

Figure 7.

At the final stage, all nodes satisfy the assertion ***rat***, and we have a fixed-point.

> *Fact* The preceding limit procedure of announcing rationality always generates the Backward Induction path.

Example, continued Backward Induction, soft scenario.

By contrast, a scenario with soft information as input does not remove nodes but it modifies the plausibility relation. Here is how we can analyze Backward Induction as a deliberation procedure forming expectations. We start with all endpoints of the game tree incomparable qua plausibility. Next, at each stage, we compare sibling nodes, using the following notion.

A move *x* for player *i dominates* its sibling *y in beliefs* if the *most plausible* end nodes reachable after *x* along any path in the whole game tree are all better for the active player than all the most plausible end nodes reachable in the game after *y*. *Rationality** (***rat****) is the assertion that no player plays a move that is dominated in beliefs. Now we perform a relation change that is like a radical upgrade ⇑***rat****: ‘If *x* dominates *y* in beliefs, we make all end nodes from *x* more plausible than those reachable from *y*, keeping the old order inside these zones’.

This changes the plausibility order, and hence the dominance pattern, so that an iteration can start. Here are the stages for this procedure in the above example, where we use the letters *x*, *y*, *z* to stand for the end nodes or histories of the game:

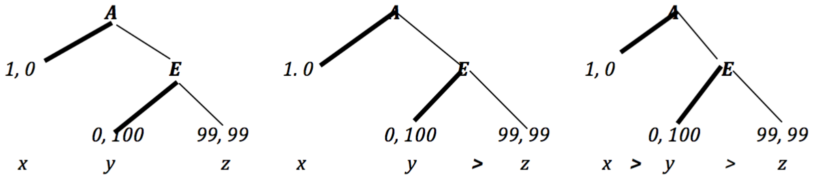

Figure 8.

In the first game tree, going right is not yet dominated in beliefs for ***A*** by going left. And so ***rat***** only has bite at ***E***'s turn, and the update makes $(0, 100)$ more plausible than $(99, 99)$. After this ordering change, however, going right has become dominated in beliefs, and a new update takes place, making ***A***'s going left most plausible.

> *Fact* On finite trees, the Backward Induction strategy is encoded in the plausibility order for end nodes created by iterated radical upgrade with rationality-in-belief.[25]

This can be proved by induction, using a natural equivalence between relational strategies as subsets of the total *move* relation and ('tree-compatible') plausibility orders on endpoints of the game tree, van Benthem and Gheerbrant, 2010. In particular, computation in our upgrade scenario for belief and plausibility and the earlier relational algorithm *BI* for Backward Induction are in harmony stage by stage:

> *Fact* For any game tree ***M*** and any *k*, $rel((\Uparrow \boldsymbol{rat}^*)^k, \boldsymbol{M})) = BI^k$.

Thus, the algorithmics of Backward Induction and its analysis in terms of forming beliefs amount to the same thing. Still, the belief limit scenario also has some interesting features of its own. One clear benefit is that it yields fine-structure for the plausibility relations that are usually treated as primitives in doxastic logic.[26]

Limit learning Limit behavior need not always stabilize in this simple way. It can get more complex with plausibility updates for other statements. Baltag and Smets, 2011 show how iterated radical upgrades for complex formulas can oscillate forever on truth of statements with beliefs occurring under dynamic modalities, though absolute beliefs will converge. Baltag et al., 2011 show how this behavior extends to 'learning in the limit' as eventual alignment of belief to the true hypothesis. They show that public announcement and radical upgrade are universal learning methods from an input stream of hard information, given a suitable prior plausibility order en-

[25] Moreover, at the end of this procedure, players have acquired *common belief in rationality*.

[26] Many of these facts can be explained by the form of the relevant statements in fixed-point logics.

coding the chosen learning method. But iterated radical upgrade is the only universal learning method under input containing a finite number of errors.[27]

Digression: social networks Beyond games, this setting applies to many other social scenarios with irreducible collective group behavior. In particular, it fits a recent trend of describing social networks (Liu, Seligman, & Girard, 2014) where agents' opinions are influenced by those of their neighbors, and long-term dynamical system behavior emerges. Such systems need not reach equilibrium in the sense of the earlier fixed-points, but they can still show other forms of stability, where the group cycles recurrently through certain patterns of epistemic states.[28]

Many earlier themes make sense here, including different zoom levels and matching logical languages. Described at one level, agents' opinions may keep oscillating, but at a level of percentages for and against, a group may be stable, as happens in equilibria studied in evolutionary game theory (Osborne & Rubinstein, 1994). These levels are entangled with the issue of automatic group behavior versus individual decisions: many unique intelligent decision makers may still add up to one statistical mob.

Returning to our focus on games, a temporal long-term perspective is also natural for players in infinite games. We can think of their strategies as compounding available individual local moves, but also at a higher zoom level, in terms of players' powers for forcing the game to produce specific sets of histories satisfying certain properties. We present one instance of a logical principle that governs the latter setting.

Example Weak Determinacy.

We noted earlier that most logic games are 'determined' in the sense that one of the players has a winning strategy. But not all games are determined (the Zermelo conditions are quite strong), and in fact, most interesting games are not. However, here is a fact from descriptive set theory that holds for players in arbitrary infinite two-player games. *Weak Determinacy* says that, either one of the players has a winning strategy, or the other player has a strategy preventing the first player from ever reaching a winning position. Now *strategies* are objects that admit of logical description (van Benthem, 2014a) as programs where conditions on actions can depend on what players know about the world or about other players.[29] But as an illustration here, we stay at a high zoom level, that of the earlier-mentioned powers of players. At that level, Weak Determinacy is a typical law of 'temporally forcing' histories:

$$\{G,i\}\varphi \vee \{G,j\}Always\neg\{G,i\}\varphi$$

[27] A more general point here is that, in this way, global limit considerations about learning methods may be brought to bear on the issue of which local update rules one should choose.

[28] van Benthem, 2015 is a first exploration of logical patterns for such long-term group behavior, ranging from generalized fixed-point logics to extended dynamic-epistemic logics. A wide variety of logics for well-known informational social group phenomena is found in Christoff, 2016.

[29] For much more on logic and strategies, see Ghosh, van Benthem, and Verbrugge, 2015.

Dynamical systems Behind all this lies a challenge. How to interface dynamic or other logics of information-driven agency with the theory of dynamical systems that underlies evolutionary game theory and formal theories of social behavior?[30]

17 Theory of play

The general picture emerging from our considerations so far may be summarized as follows. When logic, game theory and computer science come together in the way that we have sketched, something new emerges that does not belong entirely to any of these fields, which may be called a *Theory of Play*. What is in focus here is 'play' instead of 'games': how agents engage in games and game-like activities, what information they absorb, how they keep this aligned with their preferences, and what actions result in the real world. This is a huge widening of the traditional focus on game trees as such (where one makes up stories about how they could have arisen without making their ingredients explicit)—but motivations and tools for this broader enterprise can be found in all three areas mentioned.[31]

Agent diversity While Theory of Play is attractive, this agenda expansion is not unproblematic. Many issues remain in charting the great variety of logical tasks involved, and finding their static and dynamic laws.

But a more difficult further issue is how much *diversity* we allow for the agents performing these tasks: agents could have any sort of abilities for picking up information or policies for changing beliefs. To tame the resulting explosion of options in analyzing social scenarios, we need a taxonomy of natural kinds: such as risk-seeking/risk-averse agents, competitive/ cooperative, and the like. We observed earlier that most logical systems assume uniformity of agents: allowing diversity seems a real challenge to which we will return below, where we may need to relocate what is logical to the level of interfacing agents of different types.

Instead of engaging in further general soul-searching, let us end with some consequences of Theory of Play for the interface of logic and game theory. We have seen several concrete instances of this dynamic perspective already, such as our new information-dynamic scenarios for Backward Induction via iterated public announcement or plausibility upgrade. We add a few more instances, starting with a fundamental theme discussed in earlier sections.

Game equivalence revisited In an agent-oriented view, the very notion of game identity is at stake: it becomes player-dependent. It no longer makes much sense to ask, as

[30] Some first explorations of interfacing dynamic logics with dynamical systems can be found in van Benthem, 2011, van Lee, Rendsvig, and van Wijk, 2016.

[31] For instance, the players in a Theory of Play can be modeled as automata, and then an extensive literature from computational logic becomes available, cf. Grädel et al., 2002.

before, when two given games are the same: rather, we need to ask if they are equivalent *for what players*, given their preferences, beliefs, and general modus operandi. We will only give one very small illustration to show what has to change then in our earlier style of analysis of actions and powers.

Logical equivalence for rational players Let us assume that players are rational in the sense of Backward Induction. When are two games the same for players like that? It makes sense to demand that the equilibria be correlated in some way. But then, at once, earlier logical laws will fail. Consider again propositional Distribution, but now apply it to our earlier running example of a simple game with preferences.

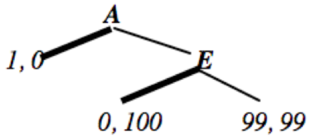

Figure 9.

The game depicted here is not Backward Induction-equivalent to its distributed form

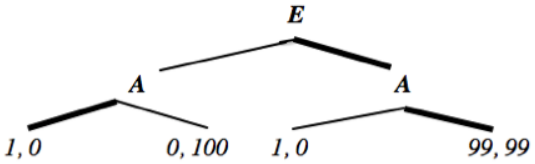

Figure 10.

as can be seen by comparing the Backward Induction strategies and outcomes in both cases. Thus, applying the logical law of distribution blindly would turn a competitive solution into a cooperative one!

It is an open problem to determine what weaker propositional algebra of game equivalence arises when we demand that the Backward Induction equilibrium outcomes are the same on both sides. For some simple, but suggestive results, see van Benthem et al., 2016.

Play equivalence But Backward Induction is only one style of reasoning-based play, and there are many others, depending on what we take the agents' modus operandi to be, perhaps including their computational and inferential limitations. The more general issue that arises then is analyzing equivalence between combinations of games plus play styles. The best way of doing this, preferably again tied to introducing matching languages, seems a serious open problem.

Bounded agency As a final theme, we mention the interaction of players that may have bounded resources of various sorts. One way of modeling these resources is in terms of 'awareness' in the recent game-theoretical literature, or of 'short sight' in the computational literature (cf. Grossi and Turrini, 2012, and the game-theoretic literature cited there). Another agent model would use various sorts of automata, and then fundamental results like the Positional Determinacy Theorem from computational logic (cf. Venema, 2015 for this and many related results) start telling us something about the reach of bounded agency.

Theory of Play has many further strands, but we leave matters here.

18 Logic, games, and general agency

In the remainder of this paper, we present a general programmatic discussion taking a still broader view. Not all intelligent interaction is game-like, and in fact, much of what we have been after in this paper has in fact been the broader arena of Logic and Agency. In the coming sections, we will show how this general shift in perspective has repercussions all around. The agency view can be taken to virtually any topic at this congress or beyond to find new issues or connect old ones.

We merely give a few examples of agency-oriented threads, many further illustrations for this perspective can be found in van Benthem, 2011, 2014a. Given the short compass of this paper, many of our claims will be somewhat apodictic, but even so, we hope that they will open some windows for the reader.

19 Epistemology: from foundation to correction

The traditional emphasis in epistemology has been on secure, cumulative knowledge claims. Even in contemporary work at the interface of logic and epistemology, finding the surplus of knowledge over belief is a focus, though there are further themes of interest to logicians (cf. Arló-Costa, Hendricks, and van Benthem, 2016; Baltag, van Benthem, and Smets, to appear).

Epistemic action In the perspective of this paper, two deviant viewpoints arose. One is that we should not focus on static attitudes like knowledge (or even belief), but in tandem with these, on the *epistemic actions* that create and modify such attitudes. Such actions involve the whole spectrum of informational events that we have discussed earlier on, from inferring to observing and communicating. But, especially in epistemology, they may well include other actions, that, say, generate doubts, or raise objections. Nothing is sacroscant.

From foundations to correction One of the most striking aspects of the agent repertoire studied here is that it does not presuppose unfailing correctness. To the contrary, mistakes and recoveries show logical ability at its best! Thus, the old foundational ideal of a safe haven once and for all for our theorizing largely disappears. It is both unreachable and not ambitious enough. *Correction* is the more exciting goal, not just correctness. Mistakes are natural, learning from mistakes is intelligent, and the real focus for epistemology is how we correct ourselves, repair our theories, and make the next creative leap to a fallible theory.

Correspondingly, our view of the role of logic changes. It is not the guardian of eternal correctness, and the key to a sterile world where nobody ever gets sick. It rather unleashes its powers in a world full of error and uncertainty, and it acts there as what one might call *the immune system of the mind*.

20 Natural language: from meaning to action

Here is an empirical angle on our program. It concerns *natural language*, the medium with which we describe the world, but also, and perhaps primarily, communicate with each other. It is illuminating to take some earlier themes to this setting.

'Of' and 'as' Recall our two directions connecting logic and games. Seeing language *as* agency, we get a dynamic view of what natural language use consists of, who uses it, and what it achieves, and all earlier topics apply. But also, looking at the language *of* agency, natural language provides a rich repertoire of expressions for driving agency, but also for discussing it, evaluating it, and reflecting on it.

However, common sense is in order. Natural language tends to blur methodological distinctions. As has often been observed, natural language is a 'universal medium' where one can discuss everything, including language itself. Our two directions not only occur in natural language, but they also allow for smooth switches. An agent can use language to communicate in a first-person 'participating stance', but also step out of this process, and comment on it in a third-person 'reporting stance'.

Natural language as agency To make this general agency perspective a bit more concrete, we quote some themes from van Benthem, 2014b.

Action/product duality Natural languages have a pervasive duality of static and dynamic vocabulary. "Dance" is both an activity verb and a noun denoting a product of that activity, and likewise for, say, "argument"—some languages do not even have a grammatical distinction here. Moreover, language has a general co-existence of static and dynamic verbs. This mirrors our approach to agency: the logic of static 'knowing' needs to be studied in tandem with the logic of the dynamic 'learning'.

Natural logic and inferential zoom A second foothold is the program of 'natural logic' from the 1980s that sought simple fast inference mechanisms that humans employ, living inside the more cumbersome full machinery of first-order or higher-order proof theory. A typical example were monotonicity inferences (performing upward or downward predicate replacement), computed by just understanding the parse trees of logical formulas (van Benthem, 1986). Inference is a family of modules, some less, some more complex—and the same might hold for our logical calculi of information dynamics. Inside their elaborate mechanisms, there may be natural language fragments that provide much simpler high-zoom reasoning about social action.

Translation as action Our final example is an agency perspective on the crucial linguistic notion of translation. Instead of a mere mapping between the syntax and semantics of two languages, translation seems correlation of behavior. This is more ambitious in that acts of communication and reasoning need to be 'translated' as well, but it is at the same time less demanding, since—in line with our earlier remarks about correction—mistakes and misunderstandings are not problematic in cross-language communication, as long as they are eventually detected and repaired.

21 Agents inside logic itself

Next, let us take agency to logic itself. Logical systems do not carry a description of their users, and presumably, these are taken to be idealized all-seeing agents that all have the same abilities. Can we tease out a parameter, and introduce agent types inside logic in a meaningful way—going to a theory of logical systems 'as used'? We need to give up hidden uniformity assumptions, finding meaningful parameters for different agents using the same system. There are some examples in the literature. Authors have looked at differences in memory, which can be modeled, for instance, with different levels in the automata hierarchy (van Benthem, 1986). Also, agents with different inferential resources have been used to model 'bounded agency', when agents may be unable to employ the full power of a standard proof system.

Fragments versus agents This perspective may change our view of many standard topics. Consider the earlier topic of 'natural logic', i.e., simple logical inference inside a complex total system. We can think of this in a standard way as a search for fragments of the full system that are decidable, or otherwise especially well-behaved. But we could equally well think that simplicity does not arise from simple fragments that guarantee success, but from simplicity of agents using complex systems. For instance, consider first-order logic as used by a finite automaton: how much correct model checking or inference could it do?

Pebble games and memory Memory modulation in standard logic occurs in 'pebble games', where access to objects in model games is restricted by a fixed supply of pebbles that are used to mark current objects of attention. As a result one parametrizes standard logical games to those that can be played successfully by players with a certain finite amount k of pebbles. However, it is typical for the state of the art that pebbling is only used in a few specialized domains, whereas it is clearly a general device for introducing memory in many logical settings. Also, players are given the same amount of pebbles, whereas again, there seems to be no need for this.

Agent diversity revisited But scenarios can be made still more exciting. What if agents are taken to be very different—as in the many scenarios of 'humans versus machines' that pervade the literature from the Turing Test onward? Our earlier logic games might now be played between different agent types, such as competing versus cooperative agents, giving up the usual uniformity of reasoners. What logical notions would then correspond to the resulting equilibria in a Theory of Play?

But this diversity also extends to other issues in logic. Consider the well-known diversity or plurality of logical systems, offering us options from classical logic to systems like intuitionistic or linear logic. What if we reinterpret this diversity, not as some sort of momentous dogmatic system choice, but as a reflection of agent variety?

These separate themes add up to a general challenge. Logic now resides, not in one ideal prescribed rationality, but at a higher level of rational interaction between different agents. How will this work precisely?

22 Logic meets reality

The preceding thoughts are still about repercussions inside theoretical academic fields. But many people read the focus on agency advocated in this paper as a step toward a more empirical account of logic as analyzing concrete reasoning and communication styles. There is certainly an influence from reality in all of the above, in that our themes are not chosen out of the blue, but in accordance with what we perceive as basic features of actual agents. This turn does not stand on its own. In recent years, logic as a discipline has been exposed to major outside influences, shaking up the cozy corners of the a priori mind. It is not the aim of this paper to chart all these influences, but the following facts are readily observable.

Logic is a source of computational devices that are transforming our world. At the same time, logic meets the empirical facts of human behavior in encounters with cognitive science. It may be hard to find logicians today who really feel at ease with splendid isolation behind the barrier of Frege's 'anti-psychologism'. Leaving the glass bead game of 'intuitions', learning instead what people really do, and how the human brain really works, is proving an irresistible challenge even to many theorists. And finally, human society itself creates an ever-growing fund of challenges such as the many whirlpools and cliffs in the world of public opinion that seems to be spinning out of control at an alarming rate.

I am not saying that our logic, games and agency view is a panacea for all the ills of our current society (or the source of all that is good), but it may become part of a reconceptualization that will equip us better to deal with major practical challenges.

In the final sections of this paper, I discuss a few themes arising in encounters with practice. I start with a simple concrete illustration, to show how the above flights of the imagination can quickly land on solid ground.

23 Designing new games and human cognition

In the real world, things go differently from theoretical scenarios. Here is an example: what happens to our standard algorithms when the world (or other agents) engage in sabotage? Rational agents should be able to cope with such changes, but let us first see what drastic effects they can have on situations we thought we understood.

From algorithms to games Under adverse circumstances computational tasks for single agents can turn quite easily into more-agent games. Consider the ubiquitous practical search problem of *Graph Reachability*:

> "Given two nodes s, t in a graph, is there a sequence of successive arrows from s to t?"

There are fast *Ptime* algorithms finding such a path if one exists (Papadimitriou, 1994). But what if there is a disturbance—a reality in travel?

Example Sabotage Games.

The following network links two European centers of logic and computation:

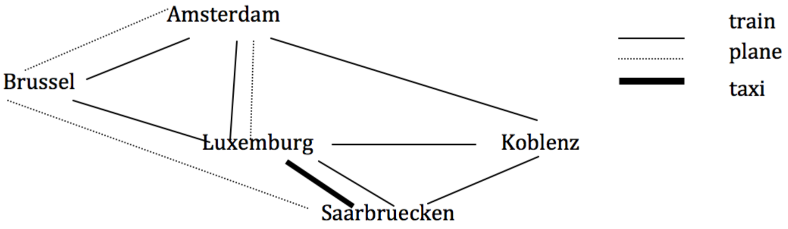

Figure 11.

Let us focus on the two nodes Amsterdam and Saarbruecken. It is easy to plan trips either way. But what if transportation breaks down, and a malevolent demon starts canceling connections, anywhere in the network? Let us say that, at each stage, the demon first takes out one connection. Now we have a two-player game, and the question is who can win it where.

Here is a Zermelo solution: the sabotage game satisfies the conditions Zermelo's theorem. From Saarbruecken to Amsterdam, a German colleague has a winning strategy. Demon's opening move may block Brussel or Koblenz, but she gets to Luxemburg in the first round, and to Amsterdam in the next. Demon may also cut a link between Amsterdam and a city in the middle—but she can then go to at least one place with two intact roads. But with a traveler starting from the Dutch side, it is the Demon who has the winning strategy. It first cuts a link between Saarbruecken and Luxemburg. If the traveler then goes to any city in the middle, Demon has time in the next rounds to cut the last intact link to Saarbruecken.

By now, sabotage scenarios have been used for various tasks, and sabotage is a general strategy for changing giving algorithms or games.[32] Here is an illustration.

Example Teaching, the grim realities.

A Student located at ***S*** in the next diagram wants to reach the *escape* ***E*** below, the Teacher wants to prevent him from getting there. Each line segment is a path that can be traveled. In each round of the game, the Teacher first cuts one connection, anywhere, the Student must then travel one link still open at his current position:

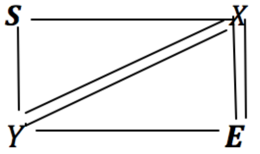

Figure 12.

Again Zermelo's Theorem applies. In this particular game, Teacher has a winning strategy: first cut a line to the right between X and ***E***, and then wait for Student's moves, cutting lines appropriately. General games like this arise on any graph with single or multiple lines, and they have been used in real teaching scenarios.

New logics of model change There is also an interesting theoretical angle here. Sabotage games suggest a language that changes the models on which it is evaluated. Thus, their logical study involves dynamic modalities referring to a new model after some change has been made to its structure, in the spirit of the dynamic update logics considered earlier, but also some genuine extensions.[33]

Gamification Like our earlier logic games, sabotage is an instance of a general phenomenon of 'gamification' that can be observed in many fields today, designing new games for pleasure or even for serious social purposes. Given the striking human penchant for interactive game play, games are a way of changing the world. At the same time, newly designed games are a free cognitive lab where we academics can observe and even manipulate how people behave in specific informational scenarios.

In all this, games constitute 'hybrid forms' of natural and designed behavior. This is one of the most intriguing features about us humans: theory can influence design qua behavior, resulting in a society where natural and designed behavior have to live together. Of course, there are many examples of this phenomenon, starting with the

[32] It can be shown that the computational complexity of solving sabotage games jumps up from *NP*-complete for graph search to *Pspace*-complete, a typical complexity class for games.

[33] Our earlier dynamic-epistemic logics are about definable model changes, and hence their complexity tends to stay low, thanks to matching recursion axioms. Sabotage dynamic logics are about arbitrary stepwise model changes, and as it happens, even sabotaged basic modal logic is already undecidable. For the latest on logics of model change, cf. Aucher, van Benthem, and Grossi, to appear.

introduction in Antiquity of specialized reasoning disciplines with specialized hybrid languages. such as mathematics or the law. It seems fair to say that we do not really understand very well how such societies function, but clearly, it will involve the agent diversity of our Theory of Play in a major way.

A universal model for cognition? We conclude this section with a grand question. We saw how computational devices become games when more users are involved, competitive or adversarial. Now consider Turing's celebrated analysis of computation, using a stylized model of a human agent performing a single task: calculation. As suggested in van Benthem, 1990, could games with agent diversity, analyzed in the same sparse style, be a universal model for human cognition?[34]

24 Logic one last time

Our final question is what the topics in this paper mean for our understanding of logic. Given the (despite all our caveats) greater empirical and practical flavor of our agency perspective, what becomes of the status of logical theory? I merely mention a few themes, leaving other valid concerns for other occasions.[35]

Normative versus descriptive? Logic and cognitive psychology or neuroscience have often been worlds apart, with logic taken to be a normative source of valid laws, while actual human behavior may fail to follow these norms for various reasons. But this dividing line seems thin. Good logical theory lets itself be informed by the best available facts about human cognition, if only to see which topics for research are most urgent or relevant. The latter point even holds if we take our logical theories of agency to be normative, uncovering the correct laws of information flow, belief revision, strategic interaction, and so on.

But the topics in this paper call for an even deeper entanglement of normative and descriptive views. Consider our emphasis on correction and learning as key aspects of agency. This was presented as crucially going beyond the statics of what is and what ought to be the case to the dynamics of *improvement*. But working toward improvement and judging something to be an improvement requires a norm, and even though these need not be absolute norms, it is a fact that learning theory is replete with norms that allow us to judge whether we are going in the right direction. So, facts and norms can and should work in tandem.

[34] The original Turing Test itself concerned a hybrid scenario with possibly different agent types.

[35] For instance, the models for bringing out dynamic actions in this paper are extensional, whereas one might think that action is a typical intensional notion, involving 'how' (that is, ways of doing things) as much as 'what' state changes are achieved. While this is true, the program outlined here for making dynamic actions explicit seems orthogonal to the choice of a level of intensionality for the logic employed.

Limitations? The high point of classical logic are its famous limitation results saying that some things are unachievable, such as decidability for logics that are strong enough, or completeness for mathematical theories that are expressive enough. It is this balance of positive and negative results that makes us logicians feel we truly understand subjects such as proof, definability or computation. But in our program so far, anything seemed to work. What would be limitations to a logical approach to games and agency? Could there be new social paradoxes that show boundaries to what the current program can achieve? I believe that there are such boundaries, and as a logician, I even fervently hope there are—but this paper has nothing substantial to offer on this score, except for the obvious fact that some logics of agency are undecidable, and can even have quite high non-arithmetical complexity.

Object and metalevel Our final point is a new philosophical worry to be resolved. We have advocated a broad view of the logic of information-driven agency, where inference is just one source of information among others, such as making observations or asking questions. So, is there no privileged role left for reasoning and its laws? One might think there still is, since at the meta-level of formulating dynamical agent systems we studied information update or belief revision in terms of their logical laws and what follows from them. So, inference would still rule the meta-level.

But if one were to be consistent with our general program, the broader informational activity perspective might also have to enter the metalevel. Perhaps the answer is this. Just as in science, we do not just *have* logical theories as sets of laws and proof rules, we also find them, develop them, and *live* them[36]—and perhaps it is that dynamic activity that constitutes the proper content of the meta-level. Theorizing about activities is itself a many-faceted activity.

25 Conclusion

The main thrust of this light programmatic paper is easy to state. To the open-minded observer, there is a great deal of logical content to agency, with games as a striking example and as a natural laboratory where theory meets practice. And looking in the converse direction, there is also a lot of agency to logic. This dual interface is a pleasure to explore, especially in the compass of games—and it suggests new perspectives on past and future interfaces of logic with a wide range of disciplines.

[36] In that spirit, the various formalisms mentioned in this paper, such as the dynamic logics of hard and soft information or various current logics of games, would not be the final laws and the last word of representation, but just one stage in formulating the meta-thory of agency.

Bibliography

Abramsky, S. (2008). Information, processes and games. In P. Adriaans, & J. van Benthem (Eds.), *Handbook of the philosophy of information* (pp. 483–549). Amsterdam, Netherlands: Elsevier.

Arló-Costa, H., Hendricks, V. F., & van Benthem, J. (Eds.). (2016). *Readings in formal epistemology*. Dordrecht: Springer.

Aucher, G., van Benthem, J., & Grossi, D. (to appear). Modal logics of sabotage revisited. *Journal of Logic and Computation*.

Baltag, A., & Smets, S. (2011). Keep changing your beliefs, aiming for the truth. *Erkenntnis*, *75*(2), 255–270.

Baltag, A., Smets, S., & Gierasimczuk, N. (2011). Belief revision as a truth-tracking process. In K. Apt (Ed.), *TARK XIII proceedings of the 13th conference on theoretical aspects of rationality and knowledge* (pp. 187–190). New York, NY: ACM.

Baltag, A., van Benthem, J., & Smets, S. (to appear). *The music of knowledge*. Netherlands: ILLC, The University of Amsterdam.

van Benthem, J. (1996). *Exploring logical dynamics*. Stanford,CA: CSLI Publications.

van Benthem, J. (2011). *Logical dynamics of information and interaction*. Cambridge, UK: Cambridge University Press.

van Benthem, J. (1986). *Essays on logical semantics*. Dordrecht, Netherlands: D. Reidel.

van Benthem, J. (1990). Computation versus play as a paradigm for cognition. *Acta Philosophica Fennica*, *49*, 236–251.

van Benthem, J. (2014a). *Logic in games*. Cambridge, MA: The MIT Press.

van Benthem, J. (2014b). Natural language and logic of agency. *Journal of Logic, Language and Information*, *23*(3), 367–382.

van Benthem, J. (2015). Oscillations, logic, and dynamical systems. In S. Ghosh, & J. Szymanik (Eds.), *The facts matter: Essays on logic and cognition in honour of Rineke Verbrugge* (pp. 9–22). London, UK: College Publications.

van Benthem, J. (2016). Tracking information. In K. Bimbó (Ed.), *J. Michael Dunn on information based logics* (pp. 363–389). Dordrecht, Netherlands: Springer.

van Benthem, J., Bezhanishvili, G., & Enqvist, S. (2016). *Instantial game logic*. Amsterdam, Netherlands: ILLC.

van Benthem, J., & Gheerbrant, A. (2010). Game solution, epistemic dynamics and fixed-point logics. *Fundamenta Informaticae*, *100*(1-4), 19–41.

van Benthem, J., & Smets, S. (2015). Dynamic logics of belief change. In J. Halpern, W. van der Hoek, & B. Kooi (Eds.), *Handbook of epistemic logic* (pp. 313–393). London, UK: College Publication.

Branderburger, B. (2014). *The language of game theory: Putting epistemics into the mathematics of games*. Singapore: World Scientific.

Carnap, R. (1928). *Der Logische Aufbau der Welt*. Berlin, Germany: Weltkreis.

Christoff, Z. (2016). *Dynamic logic of networks: Information flow and the spread of opinion* (Doctoral dissertation, ILLC, University of Amsterdam).

Gärdenfors, P., & Rott, H. (1995). Belief revision. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming* (Vol. 5: Epistemic and temporal reasoning, pp. 35–132). Oxford, UK: Oxford University Press.

Ghosh, S., van Benthem, J., & Verbrugge, R. (Eds.). (2015). *Models of strategic reasoning: Logics, games and communities*. Lecture Notes in Computer Science. Berlin, Germany: Springer.

Grädel, E., Thomas, W., & Wilke, T. (Eds.). (2002). *Automata, logics, and infinite games: A guide to current research*. Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag.

Grossi, D., & Turrini, P. (2012). Short sight in extensive games. In V. Conitzer, & M. Winikoff (Eds.), *Proceedings of the 11th international conference on autonomous agents and multiagent systems - volume 2* (pp. 805–812). AAMAS '12. Valencia, Spain: International Foundation for Autonomous Agents and Multiagent Systems.

van Lee, H., Rendsvig, R., & van Wijk, S. (2016). *Merging frameworks for information dynamics*. Copenhagen: Center for Information and Bubble Studies.

Liu, F., Seligman, J., & Girard, P. (2014). Logical dynamics of belief change in the community. *Synthese*, *191*(11), 2403–2431.

Osborne, M., & Rubinstein, A. (1994). *A course in game theory*. Cambridge, MA: The MIT Press.

Papadimitriou, C. H. (1994). *Computational complexity*. Reading MA: Addison-Wesley.

Perea, A. (2012). *Epistemic game theory: Reasoning and choice*. Cambridge, UK: Cambridge University Press.

Shoham, Y., & Leyton-Brown, K. (2009). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. New York, NY: Cambridge University Press.

Venema, Y. (2015). *Lectures on the modal μ-Calculus*. Netherlands: ILLC, University of Amsterdam.

Author biography. Johan van Benthem is University Professor, emeritus, of logic at the University of Amsterdam, Henry Waldgrave Stuart Professor of philosophy at Stanford University, and Changjiang Professor of humanities at Tsinghua University. He is a member of the Academia Europaea, the Dutch Royal Academy of Arts and Sciences, and the American Academy of Arts and Sciences. He has received the national Spinoza Award of the Dutch Organization for Scientific Research. His books include *The Logic of Time, Modal Logic and Classical Logic*, *Essays in Logical Semantics*, *Language in Action, Exploring Logical Dynamics, Modal Logic for Open Minds*, *Logical Dynamics of Information and Interaction*, and *Logic in Games*.