Logico-Computational Aspects of Rationality

Johan van Benthem, Fenrong Liu and Sonja Smets

April 8, 2019

1 Introduction

Rational behavior is a rich phenomenon, not to be captured in a single formula, but mapped out in detail in this Handbook. In what follows, we take a common sense view. We all believe, or try to believe, that our behavior is driven by reasons and reasoning, and that we are susceptible to reason, changing our minds when confronted with new facts or considerations. This is how we see ourselves, how we justify our actions to others, and how academic organizations present themselves to a general public. Further important senses of rationality, such as maximizing one's utility, will be ignored in this chapter which focuses on the main logico-computational aspects.

Logic is traditionally seen as associated with rationality since reasons and reasoning belong to the traditional province of logic. In fact, logic is often seen as rationality in its purest, and perhaps its most intimidating, form. In this article, we chart this connection in more detail without claiming that logic is all there is to rationality.

Example: valid and invalid consequence. Classical logic tells us things like this: an inference $\neg B$ from $A \to B$ and $\neg A$ is invalid (B might hold for other reasons than A) – but inferring $\neg A$ from $A \to B$ and $\neg B$ is valid, and in fact the engine of refutation. Valid inferences chained together form proofs that can yield quite surprising new insights. All this comes with a rich semantic and syntactic theory of validity which we assume the reader is familiar with.

Before we start, here is a distinction. Logical proof can be seen as a practical engine of rational behavior, and different logical systems can model rational reasoning practices. But there is also a theoretical foundational use of logic, as a study of the structure of rationality: its laws, and its limitations. In this second sense, logical analysis can target any practice that rational agents engage in: not just reasoning, but also observing, taking decisions, or debating. Both uses, practical and theoretical, will occur in what follows.

But there is a natural next step. Once we have logical systems that analyze reasoning, these cultural artefacts acquire an independent existence that interacts with human practices. In particular, reasoning is related to computational devices. For instance, the validity of the inference from $A \to B$ and $\neg B$ to $\neg A$ corresponds to a simply computable law of binary arithmetic, as can be seen in the familiar truth tables of propositional logic.

This association has proved theoretically fruitful in the foundations of mathematics, but practically, it lies behind the development of computers, information technology, and artificial intelligence that are transforming our world. Some might even say that our tools have started overtaking us.

A final part of the picture is that "us". Many themes in this article connect naturally to cognitive psychology. This interface is beyond our scope, and we refer the reader to more empirically oriented entries in this Handbook.

This article will sketch some major historical developments, identifying basic issues shaping a logico-computational understanding of rationality.

2 Mini-history of reasoning and computation

To understand the connections between logic and rationality, a historical perspective is helpful, [Kneale and Kneale, 1962], [Street, 2005]. Over time, many forms of reasoning have been analyzed and systematized in logical systems by philosophers, mathematicians, and others – a process that is still continuing. Once discovered, these systems became intellectual tools that enhance rational thinking. This invited a further step. In the early modern age, thinkers like Lull and Leibniz realized that reasoning is close to computation. From here runs a straight line to the logic machines devised by Babbage and Lovelace, and onward to modern computers and AI systems. In the workings of these machines, the notion of computation acquired sharper contours. A computing device need not be tied to one specific task, it can be programmable, the way a loom can weave different textiles depending on the current book. Thus, computing means finding algorithms for solving tasks, and since algorithms have to work on code, it also means finding data structures that represent information in appropriate ways.

All this is similar to the reality of logic itself. Textbooks say that logic is the study of inference or reasoning, but much more is involved. Reasoning presupposes a vehicle, often a language, representing the notions one reasons with. Thus, as noted in the perceptive study [Beth, 1971], logic has always been about a tandem of proof and definition, or if you wish, proof theory and model theory. And, Beth added that a crucial third constant of logical thought was the notion of algorithm, which combines the former two.

As remarked in [Minsky, 1961], the history of computing has a curious character in that its major principles were discovered before its actual practical success, a situation unlike that in many other disciplines. In the 1930s Gödel analyzed the limits of what logical proof systems can achieve, finding that logical systems whose expressive power suffices for encoding basic arithmetic are either inconsistent or incomplete: unable to prove all intuitive mathematical truths about their domain [Gödel, 1931]. Gödel's proof involved a deep analysis of computable functions (the 'recursive functions'), which was taken further by Turing, who gave concrete machines that can compute all recursive functions [Turing, 1936]. Indeed, there is a universal Turing machine that, given any program code and input, computes the effects of running that program on that input. In this setting, Church showed that standard systems such as first-order logic, though completely axiomatizable, are undecidable: no computing method can decide, for arbitrary first-order consequence problems, whether or not they are valid [Church, 1936]. Thus a major trade-off came into view: increased expressivity of a language and complexity of the decision problem for validity are at odds eventually.

This history highlights three points that seem crucial to understanding the nature and scope of rationality even today. The first may be considered a practical issue of modus operandi. If rationality has a computational engine, how should we understand its tandem of reasoning and concept formation? The second point is theoretical. Are there principled limitations to logic-based rationality, say, in terms of natural tasks that lie beyond the scope of rational inquiry? Gödel's theorems keep generating discussion, [Smullyan, 1994], [Wang, 1996], and a common idea is that there is more to rational thinking than what is captured in logical systems. Even so, when this "more" can be explained further, and spelled out in terms of computation, the great limitation theorems of logic apply again. A third point is again practical. The historical results on what proof systems and computing devices cannot do were in fact immensely helpful in the further development of systems of inference and computation that can do a lot. Likewise, the modern 'challenges to rationality' discussed later in this article may provide an impetus for deeper insights into what rationality can achieve. Having said that, much of the modern literature in AI or cognitive psychology is of the 'can do' type: one seldom reads about the discovery of deep new limitation theorems.

The foundational era of Gödel and Turing may be seen as telling us what is computable or provable in principle, or not. While this high abstraction level for viewing information remains a valid perspective on rationality, the subsequent history has yielded many further themes that are of relevance.

3 Computer science and artificial intelligence

Computer science. The development of computing in the 20th century has generated ever increasing practical achievements, but also an ever growing insight into fine-structure. There are different models for computation, from Turing machines to many other devices, and crucially, these models come in hierarchies. Some tasks are solvable by simple finite automata, other require memory management to varying degrees. Likewise, there is a wide diversity of, poorer or richer, languages for specifying data structures and writing programs, [Harel, 1987]. And eventually, around 1980, the whole idea of computing architecture moved away from single Turing machines to networks for distributed computing, the reality of computing today [Andrews, 2000]. This fine-structure has given rise to new mathematical fields such as automata theory [Chakraborty et al., 2011], complexity theory [Papadimitriou, 1994] and process algebra [Bergstra et al., 2001], that chart the varieties of computation in different ways, many of them still connected to logic. This historical process is still ongoing, and the foundations of computation remain under debate. For instance, there is no consensus yet on a definitve notion of algorithm, a more intensional notion than the extensional input-output behavior generated by Turing machines [Haugeland, 1997], [Bonizzoni et al., 2013].

This development comes with notions and insights that are relevant to understand rationality. The fine-structure of computation gives a precise meaning to the earlier-mentioned double aspect of modus operandi: reasoning engine and representational apparatus. And the variety encountered in actual computation suggests that 'boundedness' of resources and powers is the norm in rational task performance, as opposed to having one idealized super-device. Matching this practical concern is a fundamental issue. At the level of the complexity theory of space and time resources needed for performing computational tasks, we are really talking about information that is available in the world and how to process it. But this forces us to think what is the information available to rational agents [Adriaans and van Benthem, 2008]. And there is yet more to be learnt from the world of computing. If we think of rational agents as being able to perform many tasks, just as computers and networks of computers can, then there is a fundamental issue of architecture. How do the different components of the overall system pass information and cooperate (see e.g. [Gabbay, 1998])?

Artificial intelligence. Moving closer to humans, computer science flows over seamlessly into AI. From the start, computers have been seen as a powerful model for human intelligence. In an interesting departure from the detailed internal analysis of computing by Turing machines, the famous Turing Test

approaches intelligence in the classical tradition of measuring theoretical notions by external observable behavior, [Turing, 1950]. It proposes that a computer achieves human intelligence if an observer communicating in natural language cannot tell that computer apart from a human by asking questions and engaging in conversation. Over the years, computers started to pass variants of the Turing Test, or other types of intelligent behavior. Interestingly, none of these are usually considered conclusive, as the criteria are a moving target, [van Harmelen et al., 2008]. Passing the test is dismissed as not a display of 'real intelligence', and then the demands are shifted a little further. But behavioral tests are crucial to judging human rationality as well. We seldom look inside people's heads to monitor their considerations, but observe their words and actions.

A final intriguing feature of the Turing Test is its hybrid scenario where different types of agents, humans and machines, interact, presaging the reality of human-machine interactions in modern society. This scenario goes beyond the classical 'emulation' or 'competition' concern: how can societies of mixed agents, with different strengths and weaknesses, interact successfully (see [Wooldridge, 2009], [van Benthem, 2014])? The resulting perspective of diversity in agency is only beginning to make itself felt more widely. Most paradigms in logic, or philosophy, assume that agents have similar abilities for reasoning and information processing, though their information and preferences may differ. Such uniformity assumptions underlie most theorizing about generic notions like 'humans', 'rational actors', and so on, and they may even seem to embody moral imperatives, like in treating everyone equally qua rights and duties in ethics. If one accepts diversity, however, a host of issues, notions, and theory concerning rationality will have to be rethought.

While these trends in AI and computer science extend the agenda for logical conceptions of rationality, recent challenges call the very logic-based approach into question. We will discuss two instances in Section 6: the 'high' versus 'low' rationality competition in understanding social agency, [Skyrms, 2010], and the rise of non-representational machine learning techniques, [Kelleher, et al., 2015]. But to keep things in historical perspective, for now, here is another major trend from the 1980s onward which has turned out relevant to the study of rationality, bringing together ideas from the worlds of computation and philosophical logic, [Gabbay and Guenthner,1983-], [Gabbay, et al.,1993], [van Benthem, 2018]. After explaining what is involved in Section 4, most of the concrete examples of this article will be found in Section 5, showing logic at work in this modern setting.

4 From machines to agents

Human agents have a much wider range of rational activities than just reasoning from given information, elucidating what was already implicitly there. The information flow that guides action is dynamic. Agents constantly pick up new information from their environment through observation and communication, and they can search their memory for old information, too. Rationality is about picking up relevant information as much as reasoning, as is amply demonstrated in both daily life and the history of the sciences.

However, even processing of reliable information, no matter how rich, is just one dimension of rational agency. Information can be less or more reliable, and agents do not just accumulate knowledge, but also form beliefs that can be shown wrong by new information. Thus, rational agents are not those who are always correct, but those who learn from errors, revise their beliefs, and in general, have the capacity to correct themselves, [Popper, 1963], [Kelly, 1996]. Robust rationality shows in dealing with situations where one is proven wrong, rather than those where one is right. Logic still has a role to play here, since many models of belief revision and learning come from logic [Baltag and Smets, 2008]. Belief revision and learning are dealt with in article [van Ditmarsch, 2019] of this Handbook.

And rational agency does not even stop here. Many notions of rationality are not purely informational, but are about maintaining a harmony between an agent's information and beliefs on the one hand, and on a par, the agent's preferences, goals, or intentions. One can discuss which connection is essential here, whether maximizing expected utility or some other bridge law, and agents may differ in this respect, but the point is the balance of information, goals and actions maintained by truly rational agents.

Summing up, a rational agent can gather information in a variety of ways, integrating observation, inference, and communication. In this process, the agent can form a rich variety of attitudes, ranging from knowledge and belief to disbelief or doubt, [van Benthem, 2011]. Moreover, a rational agent can function in an uncertain environment where beliefs can be shown wrong, and can learn from errors over time. And all these things maintain a purpose, a balance between the agents' goals and the information and actions required. What that precise balance is will depend on one's particular view of rationality, there is no need to assume total uniformity. And even this is not yet the full picture. Rational agents are able to display their skills in social interaction, a topic we will turn to later on.

This richer notion of rational agency has long a concern of philosophers and philosophical logicians, witness the various chapters of the *Handbook of Philosophical Logic* ([Gabbay and Guenthner,1983-]). In the 1980s, a similar

move to a richer picture of agency emerged in computer science and AI, in the area of multi-agent systems [Fagin et al., 1995], [Shoham and Leyton-Brown, 2008], [Wooldridge, 2010]. Its driving force is a shift in thinking about computing devices, from machines to agents with a behavior best understood in terms of basic features we normally ascribe to rational humans.

A further parallel is the rise of autonomous systems in AI. Robots can pick up information from their environment through sensors, decide and act in performing their tasks, [Cardon and Itmi, 2016], [Brafman et al., 1997], [Su and Sattar, 2008]. Thus, robots display a spectrum of information-processing and goal-fulfilling capabilities that gets closer to real human agents as described above. This is not a one-way street. For instance, as real-life robot sensors are only accurate up to some margin of error, epistemic aspects come into play for their goal specifications, [Brafman et al., 1997], and this acting on evidence of various qualities has already inspired work on new models for evidence-based belief in epistemology, [van Benthem and Pacuit, 2011].

There is a natural transition here to richer mathematical models and a further field. Agents with different powers for observing information and choosing actions, pursuing goals of their own, are like players of games. Accordingly, computer science and game theory have started drawing closer, [Nisan et al., 2007], and logic, with its connections to games of argumentation and information-seeking, [Hintikka, 1973], is a natural partner, [van Benthem, 2014]. Indeed, the new area of epistemic game theory may be seen as a venue created by these contacts, cf. the chapter [Perea, 2019] in this Handbook.

Even so, we do not have a canonical view of what a rational agent is and does similar in elegance and fertility to that of a computing machine, let alone a model for a universal rational agent comparable qua sweep to the universal Turing machine. In fact, the drive toward diversity noticed earlier may make us think about different kinds of rational agents, rather than uniqueness, changing our assessment of performance. For instance, is a rational agent someone who wields enormous cognitive powers, or someone doing their best with limited powers? Or in a multi-agent setting, are rational agents those who perform well against other rational agents, or those who do well with a large bandwidth of different types of agents in their environment?

5 Logical models of rational agency

In recent decades, many features of rational agents have been studied by logicians, extending the agenda of the field while retaining standard methods. Information update and knowledge change are studied extensively in a variety of temporal logics of agency, [Parikh and Ramanujam, 2003], [Belnap et al.,

2001], [Fagin et al., 1995]. Another broad approach is dynamic-epistemic logic, [Baltag et al., 1998], [van Benthem, 2011], [van Ditmarsch et al., 2007], whose semantics models the dynamic processes whereby agents form and modify representations of the information at their disposal, something that is not inference, but that can be described just as precisely in logical terms.

Example: dynamic logic of information flow. In a simple two-party dialogue, Agent 1, who is uncertain about the truth or falsity of p, asks "p?". A second agent then truthfully and publicly replies "yes". Analyzing the information flow in the dialogue, the first agent's question conveys that she doesn't know the answer but also that she thinks the second agent, a fully reliable source, does know. The second agent's answer conveys the information that 1's assumption about 2's knowledge was correct, and moreover after 2's answer, p is common knowledge in the group of these two agents. More complicated examples, such as the famous Muddy Children puzzle [Fagin et al., 1995], illustrate how truthful public announcements of uncertainty to a posed question can gradually lead the agents to knowledge. The symbolic language capturing all of this can make assertions of knowledge change after a public announcements via statements of the form $[!\varphi]K\psi$, which capture that the agent will know ψ after the public announcement of assertion φ. A key principle of information flow in the logic of public announcements is the recursive equation stated in the equivalence: $[!\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi)$ which relates the expression $[!\varphi]K\psi$, capturing the knowledge after the event $!\varphi$ happened, to its 'pre-encoding' via conditional knowledge that the agent had before the event happened. Such principles illustrate the interchange mechanisms between dynamic operators for informational events and epistemic operators expressing standard attitudes of agents, a crucial ingredient in our logical encoding of information update and knowledge change.

Logics of belief change under hard and soft information have been developed in the same style, [van Benthem and Smets, 2015] though there are also other formal paradigms with the same purpose, see the entry by H. Rott in this Handbook on AGM theory. The notion of learning fits naturally with belief revision, and connections between dynamic logics of knowledge and belief with formal learning theory can be found in [Baltag et al., 2011].

Example: Belief change. Referring back to the above two-party dialogue, now assume that Agent 2 is not a fully reliable information source. Starting from the initial situation in which Agent 1 believes neither p nor $\neg p$, the answer of 2 to her question can trigger 1 to change her mind and possibly to adopt a wrong belief. Yet how exactly she changes her mind will depend on the level of trust that 1 has in 2 as an information source about p. If we focus on the case in which 2's answer "yes" is considered to be reliable but not infallible,

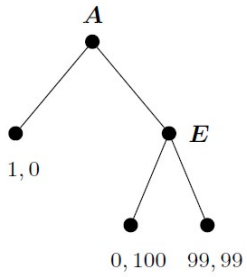
the 'belief upgrade' that it triggers can be more radical and induce a strong belief in p, or more conservative, inducing a weak belief in p. Again, this process obeys logical laws. In a logical language we add constructs of the form $[\uparrow \varphi]$ for conservative belief upgrades and $[\Uparrow \varphi]$ for radical upgrades. This brings to light many laws – for instance, $\neg K \neg P \rightarrow [\uparrow P] BP$ says the agent comes to believe that P is the case (more precisely, that P was the case before the announcement), unless she already knew (before the announcement) that P was false. In logical studies of learning, one studies iterations of such upgrades and analyzes how well they perform as a learning method.

The study of key features of rationality in this style continues. Further aspects of purposeful rational behavior brought into the scope of logic include the management of current issues that guide inquiry for particular tasks at hand, cf. the inquisitive logics of [Ciardelli and Roelofsen, 2011].

Next, moving from informational tasks to agents' preferences and goals, which determine how they evaluate situations, [Liu, 2011] studies dynamic logics of preference change, while [Harrenstein, et al, 2001] studies goal dynamics. Preference dynamics shows similarities with deontic logics describing what is obligatory and permitted for agents in environments where new commands can change moral ordering of situations and actions, [Yamada, 2008].

As said above, a truly rational agent shows a balance between information and preferences or goals. Logics combining all these features occur in influential frameworks in the field of multi-agent systems such as BDI [Rao and Georgeff, 1991], inspired by [Bratman,1987], describing how agency is driven by a balance between beliefs, desires, and intentions. But perhaps the most active research area where action, information about the world and others meet is in the logical study of strategic behavior and equilibria in games.

Example: reasoning about extensive games. Consider a finite game with two players A and E, and outcome-values written in the order (A-value, E-value):



Intuitively the outcome (99, 99) seems best for both players but that is not what the standard game solution method of backward induction yields.

Looking from the bottom to the top, if E is to play she will choose left, and so, if A believes that E will make this choice, she herself will play left in the first round and end the game, with outcome (1, 0). Analyzing why players should act in this way, involves the interplay of many notions including players' actions, beliefs, preferences, beliefs, and plans, all of them long studied by logicians, be it usually in different settings or as separate topics. For precise logical definitions of equilibrium outcomes in games, and many concrete examples of logical game analysis, cf. [van Benthem and Klein, 2019].

All the logics mentioned here exemplify the earlier-mentioned tandem of algorithm and data lying at the heart of computation. As can be seen in our examples, the dynamic process with its various events that produce new information or new goals operates on appropriate static models that support standard attitudes of knowledge, belief, preference, and the like.

Digression. There are also approaches folding all of the above activities under varieties of inference, emphasizing departures from classical consequence to non-classical non-monotonic logics and resource-conscious substructural logics [Restall, 2000], [Restall, 2005], [Horty, 2014]. For a discussion of these two methodologies and their interconnections, cf. [van Benthem, 2018].

Example: non-monotonic logic. Monotonicity is the property of valid consequence in classical logic that, if $\Gamma \models \varphi$ and $\Gamma \subseteq \Gamma'$, then $\Gamma' \models \varphi$. This fails for defeasible or default reasoning. Say, if I know that Tweety is a bird, I can conclude that Tweety flies, yet with further information that Tweety is a penguin, the flying no longer follows. The field of non-monotonic logic studies properties of this new setting. In contrast, dynamic logics of belief revision in the above style capture default phenomena on a classical base, locating the non-monotonicity in belief change rather than in inference rules.

The logical study of ever more aspects of agency aims at a non-purely behavioral view of rationality and intelligence by identifying key internal features and mechanisms. But combining logical and computational agendas does not make logical systems realistic software agents or human agents. Their idealizations are far from the concrete data needed for algorithms to work, and the semantic slant of most logics of agency sits uneasily with implementation, which needs syntax. Recent studies try to mediate between semantic models and syntactic representation for computing agents, [Lorini, 2018], [Swayamdipta et al., 2018], [Halpern and Rego, 2009].

Also, the development of ever richer models raises questions. Where is the boundary of agency, as more and more topics are taken on board, and what is 'rational' about the activities so described? And in all this, are we now describing what agents do, or are these logical systems still normative?

A common view is that all logical systems mentioned describe idealized laws that may or may not be followed by actual agents. And this tension may be just what is needed. It makes no sense to say, for instance, that belief revision leads to 'correction' of earlier beliefs unless we have a norm for what we consider to be 'correct' in the given circumstances.

6 Rationality in social settings

Interactive scenarios In the modern study of agency, an important change has occurred reflecting the fact that distributed systems and networks are the main engine of computing today, not single machines. Likewise, in multi-agent systems, [Wooldridge, 2009], [Shoham and Leyton-Brown, 2008], we are usually not dealing with single agents, but with groups of interacting agents, sometimes even crowds or societies. This shifts the location of intelligence and rationality from just what individual agents do and are capable of to the quality of their interactions. And potentially, it shifts the locus from what individuals want and do to emergent properties of the social system. We will discuss this trend a bit further, including its challenges.

The interactive perspective is not alien to logic. Ever since Antiquity, dialogue, argumentation and debate have been paradigmatic scenarios, and as we have noted already, there is a rich interface of logic and games, [Hodges, 2018], [van Benthem, 2014]. A core feature here, studied in epistemic logic, is the ability of rational agents to represent and reason about others, leading to iterated knowledge of the form "agent i knows that agent j knows that" and the like, and its counterparts for belief and other attitudes. This recursion to higher levels is everywhere, positive or negative: we can even be afraid of fear, of fear of fear, and so on. The actual extent to which human agents can display these abilities is studied in cognitive psychology under the heading of Theory of Mind, [Premack and Woodruff, 1978], [Isaac et al, 2014]. Iterated knowledge is also used widely in computer science for the analysis of correctness and security of communication protocols, [Fagin et al., 1995].

But there is much more to social interaction than epistemic reflection. Strategic action involves complex dependencies of one agent's behavior on that of other, or perhaps better, his expectations about the behavior of others, [Aumann, 1995]. Here is a simple illustration of what happens to computational tasks when they have to be performed in an interactive setting.

Example: sabotage game. A Traveler in a graph uses links to travel from one node to another in order to reach some specified goal region, yet she faces a malevolent Demon who cancels a link after each move she makes.

After that, Traveler can go along some still existing link, and so on. This 'sabotage game' models search tasks under adverse circumstances and other basic scenarios. The problem of graph reachability for a single player is in Ptime, but the solution complexity of the sabotage game is Pspace-complete. Logic helps determine who has a winning strategy in given sabotage games by defining the essential challenge-response pattern, and it helps reason about general properties of such games, [van Benthem, 2014]. This is just one case where logic meets 'gamifications' of agency scenarios, and in fact, logics are even used to devise concrete new practical games, becoming tools of design as much as of analysis, [van Benthem and Klein, 2019].

With preference added, various notions of rationality have been investigated by logical means, such as those embodied in game solution methods like Backward Induction, Iterated Removal of Strictly Dominated Strategies, [Osborne and Rubinstein, 1994] or Iterated Regret Minimization, [Halpern and Pass, 2009]. The structure of strategies by themselves is being studied extensively at the interface of game theory, logic, and computer science, [Brandenburger, 2014], [van Benthem et al., 2015]. Also, cross-over research has sprung up between these fields, witness the Boolean games of [Harrenstein, et al, 2001], where players can each manipulate the truth value of propositions toward achieving their goals.

However, not all is plain sailing here, and various challenges emerge.

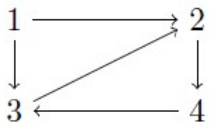
High and low rationality. The interface with games also poses a challenge to the analysis of rationality. Game theory comes in two flavors. Classical game theory has rich agents that deliberate and design complex strategies, [Osborne and Rubinstein, 1994], evolutionary game theory can make do with poor agents, perhaps hard-wired biological types, [Maynard Smith, 1982].

Example: evolutionary games. In a 'Hawk-Dove' game, two individuals compete for a resource and can adopt either a Hawk or a Dove strategy. Hawks fight aggressively against other Hawks, in order to obtain the resource, until injury occurs and one retreats, or just takes the resource from a Dove, while a Dove retreats when facing a Hawk, while two Doves share the resource. Game theory computes equilibria here, which are typically in mixed strategies. With repeated Hawk-Dove games, the appropriate notion is that of an 'evolutionarily stable' strategy, and it can be show that certain mixtures of Hawk-Dove populations are stable, when the value of obtaining the resource is greater than the cost associated with possible injury in a fight. One can think of these mixed strategies as complex behavior for individual reasoning agents, but also as percentages of a population consisting of two types of agent, each just doing what it does, perhaps for biological reasons.

In the term of [Skyrms, 2010], the realm of complex reasoning players is that of 'high rationality', the realm of hard-wired simple agents that of 'low rationality'. Often the latter seems to do as well as the former. For instance, a classical game-theoretic analysis might say that through some Kantian or Rawlsean argument involving thinking about others, we all arrive at the conclusion that we should live by the principles of morality, with the exception perhaps of a few ungifted reasoners. By contrast, an evolutionary stability argument in evolutionary game theory may tell us that in the long run, a population of simple law-abiders (the prey) and law-breakers (the predators), who both cannot help being what they are, is stable. No reasoning need be involved at all, the morality is emergent system behavior. This influx from evolutionary game theory reinforces the message of distributed computing: a society of many simple agents can produce highly complex behavior.

Here is one more case of emergent long-term complex societal behavior.

Example: limit behavior in social networks. Consider the following simple network, with a modal formula $\Box p$ saying that p is currently true at all neighboring nodes. The update rule is given by $p = \Box p$, which means that agents follow what all their neighbors do. It will be applied iteratively:



Runs of this system can easily be computed:

Case 1: Initial $p = \{1\}$. The second stage has $p = \emptyset$, and this remains the outcome ever after.

Case 2: Initial $p = \{2\}$. The next successive stages are $\{3\}$, $\{4\}$, $\{2\}$, and from this stage onward, we loop.

Case 3: Initial $p = \{1, 2\}$. The next stage is $\{3\}$, and we get an oscillation as before in Case 2.

Case 4: Initial $p = \{1, 2, 3\}$. We get $\{1, 3, 4\}$, $\{2, 4\}$, $\{2, 3\}$, $\{1, 3, 4\}$, and an oscillation starts here.

Thus, we see how network update dynamics can stabilize in one single state (witness Case 1), but also oscillate in loops of successive predicates. These oscillations come in different forms. Sometimes, successive models in the loop are very similar, in fact isomorphic (Cases 2 and 3), sometimes the loop runs through different non-isomorphic network configurations (Case 4).

To put the challenge starkly, perhaps logic-based rationality is *not necessary* to understand the social behavior of human and artificial agents. But things are more complicated. In daily life, we think carefully in the 'high' style about certain issues, but given our limited resources, we just follow, 'low'-style, our neighbors on perhaps the majority of issues. This mixture calls for explanation, and current investigations are charting its details.

Combined scenarios. Here are a few examples. [Liu, et al., 2014] study a simplest action of an agent in a social network, viz. following one's neighbors' preference, belief or behavior via some rule: say, following the majority, or above some threshold. This process, called 'diffusion' in sociology, models spread of cultural elements and new ideas. An agent's epistemic states and its dynamical changes can be modeled by automata and their state transitions, where transition rules can vary between agents, reflecting their different types. Taking a cue from [Skyrms, 2010], a logical characterization can be given of conditions for stabilization of agent's beliefs: so, long term system behavior can be predicted and computed. This framework combines ideas from sociology, [French, 1956] [Friedkin, 1998], with epistemic logic, adding an essential element to logical models: the structure of the social network. Repercussions for the interplay of individual and social epistemology are explored in [Shi, 2018]. For networks composed of high rationality agents, [Baltag, et al., 2018] investigates the interplay of epistemic and social information flow, showing, for instance, how an individual agent's knowledge about distant neighbors allows them to anticipate behavioral changes in their immediate environment. Further work in this social-epistemic line is found in [Seligman et al., 2013], [Xue, 2017], [Christoff, et al., 2016], and [Smets and Velazquez-Quesada, 2017].

For another example, consider a group of individually rational agents who reason collectively towards a common decision. Two scenarios can happen. Individual agents can enhance each other's reasoning power and bring about a higher level of group rationality surpassing the abilities of each individual agent. But the opposite can also happen: agents may find themselves locked in a social scenario that leads to irrational group behavior. Whether the one scenario will happen or the other is investigated in [Baltag et al., 2018], in terms of differences in interests and abilities between agents. One striking conclusion is that 'irrational' group behavior is often not caused by irrational behavior of individual agents, but by the misalignments of their interests.

Social herding phenomena yield another contrast between individual and group rationality. In an informational cascade, [Bikhchandani et al., 1992], a sequence of individual agents follows the decisions of their predecessors, while ignoring their own private evidence. In [Baltag, et al., 2013], the question

is addressed whether individual rational agents, who use all their higher-order reasoning power, can stop a cascade from happening. The answer is surprisingly 'no', and this fact can be proved by logical techniques that track information updates. However, the protocol matters. When agents have total communication and sharing of evidence, cascades can be stopped. These protocols can be studied as strategies, leading to connections with he earlier-mentioned logics of games.

We have elaborated on the above research to show how ideological differences between high and low rationality turn into a careful study of how the two interface. In achieving that, there is a methodological issue of connecting two kinds of mathematics: logic, and the dynamical systems theory underlying evolutionary game theory and social network theory. Recent explorations in this direction are [Kremer and Mints, 2005], [van Benthem, 2015], [Klein and Rendsvig, 2017], [Hornischer, 2019].

7 A changing world

The current world of computing and AI offers several challenges to logic-based paradigms of rationality. In Section 6, we already discussed one of the intriguing fundamental issues to come out of this: the interplay of high and low rationality. In this section, we briefly discuss two more.

Probability. One conspicuous feature, both in practice and in theory, is the extensive use of probabilistic methods, a quantitative paradigm often seen as being at odds with qualitative logical analysis. Probability underlies many modern computational systems, it lies at the heart of game theory and dynamical systems theory, and even in epistemology, probabilistic styles of analysis are at least as widespread as logic-based ones.

The challenge here is not one of replacement, but of combination. Qualitative and quantitative approaches co-exist in many areas, and their connections are a matter of continuing investigation. For instance, uncertainty, both ontic and epistemic, and our ability to reason about it, is a key aspect of rationality. Epistemic and doxastic logics, static and dynamic, model uncertainty in terms of ranges of options, [Adriaans and van Benthem, 2008], whereas say Bayesian epistemology uses updates of probability functions, [Talbott, 2016]. The compatibility of the two perspectives shows in combined probabilistic logical systems, [van Benthem, 2009], [Halpern, 2005], designed to reason about different types of ontic and epistemic uncertainty, and bringing together the logic-based updates of Section 5 with Bayesian conditioning and Jeffrey conditioning. But there are also deeper connections. The foundations

of probability were close to logic, as is shown in the work of Boole and De Finetti, while various strands of recent research link the two realms in new ways (see e.g. [van Lambalgen, 1996]). The authors in [Harrison, et al., 2018] study low-complexity qualitative reasoning systems that admit of introducing probability measures, while [Leitgeb, 2017] studies the opposite direction: deriving qualitative notions of belief from richer probabilistic models.

Other uses of probability are about action rather than information, witness the importance of probabilistic mixed strategies enriching the space of possible behaviors in game theory and dynamical systems. As we have said, interfacing with logic seems possible, though non-trivial, [van Benthem and Klein, 2019]. But there are also quite different interfaces of logic and probability, for instance in the innovative DOP architecture of [Bonnema, et al., 1999] [Bod, 2008] which combines classical rule-based models with pattern recognition in a probabilistic memory of earlier performance in natural language processing, as well as in other forms of expert behavior.

There are many further philosophical and technical issues to be explored at this rich and growing set of interfaces that we cannot cover in this article (see e.g. [Spohn, 1988], [Spohn, 2012]).

Machine learning. A second major use of probabilistic and other quantitative methods occurs in what may be the highlight of current AI: machine learning, [Kelleher, et al., 2015]. These methods work well on large sets of data, outperforming symbolic approaches that tend to have problems of scalability. As just one example, in supervised learning, a neural network is constructed consisting of nodes with adjustable thresholds and links between nodes of adjustable strengths. Each current setting for all of these produces an activity in the output layer given an input to the initial layer of the network. A cost function measures the distance of the current outputs to the desired ones on the training inputs. The network can then adjust its weights and thresholds in the direction of lowering the cost function by well-known mathematical techniques such as gradient descent, [Russel, et al., 1994]. In the end, stable optima in network activation are reached that turn out to work amazingly well in new cases outside of the training set, at least in many important computational and cognitive tasks. These networks, related to so-called spin glass models in physics [Nishimori, 2001], and indeed this whole methodology for machine learning reflects general statistics for any sort of setting, rather than specifically human features. Networks like this are also reminiscent of the social networks of Section 6, though now made more dynamic and learning-oriented.

Neural networksqs in machine learning do not have anything obvious corresponding to classical logical models. There is no language and no rep-

resentation, and the dynamic operations of the network do not reflect logical operations in any obvious manner. Also, very different stable states of the network resulting from training sessions can perform the same tasks, and the invariances are hard to detect. Thus we arrive at what looks like a second major challenge to logical methodology. Whereas low-rationality methods raised the question whether logical analysis was *necessary*, deep learning methods raise the question whether logical analysis is *possible*.

It is far too early to adjudicate this particular debate, but the fact that there is no obvious classical analysis pattern to machine learning systems does not mean that none can be found. Indeed, there are some interesting developments tending toward cooperation rather than animosity. Integrating statistical inference in neural networks and learning with symbolic programs and symbolic reasoning is an active area of research, [d'Avila et al, 2015], [Baggio et al., 2015], [Leitgeb, 2004], and [Balkenius and Gärdenfors, 2016], and there are also more recent strands. There is an ongoing line of work on 'explainable AI' trying to find humanly intelligible qualitative patterns behind opaque quantitative systems, with a growing body of research into causal structure and causal reasoning, [Pearl, 2000], [Halpern, 2016], [van Rooij and Schulz, 2019], information flow from dependencies, [Baltag, 2016], [Väänänen, 2007], [Sandu, 2012], and axioms from conditional logic as a way of classifying types of machine learning [Ibeling and Icard, 2018].

Finally, a distinction made at the start of this article should be kept in mind with all challenges that we have mentioned. If logic is only seen as a practical ground-level account of rational activities of reasoning, observation, and the like: then other frameworks look like competitors. At best, the question might then become whether logic can still enhance such methods in terms of representation or computation. But in the foundational sense of logic as an analysis of the structure of theories of computation and agency, even probability theory, dynamical systems and machine learning have logical structure, and a meeting of the minds seems entirely feasible.

8 Conclusion

This article has presented some broad perspectives from logic and computation on rational agency. These ranged from high-level foundational insights into information and proof to specific studies of various abilities of information- and goal-driven agents. A rational person, in this light, is a reasoner, information processor, concept crafter, and purpose seeker: fallible, but talented. Is this sort of agent a real cognitive agent? We have left the matter of cognitive reality aside, and defer to other articles in this Hand-

book, for instance the entry by *van Lambalgen & Stenning* connecting logical and computational tools to cognitive (neuro-)science.

The main thrust of a logical approach as we see it is theoretical, but the deep entanglement of logic and computation over the last century sketched here has added practical dimensions. Rationality as studied here can be programmed and put into intelligent systems, even though the path to feasibility is not easy or trivial. It is this very distance that allows logical theories to be (more) normative, providing an essential tension between the real and the ideal in the study of rational behavior that sparks further investigation.

We have not hidden the fact that the classical logico-computational paradigm faces challenges, coming from probability theory, dynamical systems, and machine learning. But we think this is all to the good, since these challenges suggest new interface topics of interest to all.

Finally, it should be clear that we have not claimed that logic is the only game in town. Neither is computation. The approach surveyed in this article does not hold the unique key to understanding the rich phenomenon of rationality, but it does offer one valid and illuminating perspective.

References

[Adriaans and van Benthem, 2008] Adriaans, P. and van Benthem, J. (2008). *Handbook of the Philosophy of Science, Vol.8, Philosophy of Information*. Elsevier Science B.V.

[Andrews, 2000] Andrews, G. R. (2000). *Foundations of Multithreaded, Parallel, and Distributed Programming*. Addison–Wesley.

[d'Avila et al, 2015] d'Avila Garcez, A., Besold, T., de Raedt, L., Földiak, P., Hitzler, P., Icard, T., Kühnberger, K., Lamb, L., Miikkulainen, R., Silver, D. (2015). Neural-Symbolic Learning and Reasoning: Contributions and Challenges. AAAI Spring Symposium - Knowledge Representation and Reasoning: Integrating Symbolic and Neural Approaches: Stanford University, Palo Alto, CA

[Aumann, 1995] Aumann, R. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8(1), pp. 6–9.

[Baggio et al., 2015] Baggio, G., van Lambalgen, M., and Hagoort, P. (2015). Logic as Marr's computational level: Four case studies. *Topics in Cognitive Science* vol. 7(2):287–298.

[Balkenius and Gärdenfors, 2016] Balkenius, C. and Gärdenfors, P. (2016). Spaces in the Brain: From neurons to meanings. *Frontiers of Psychology*, 22 November 2016.

[Baltag et al., 2011] Baltag, A., Gierasimczuk, N., and Smets, S. (2011). Belief revision as a truth-tracking process. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK XIII, pages 187–190, New York, ACM.

[Baltag et al., 1998] Baltag, A., Moss, L., and Solecki, S. (1998). The logic of common knowledge, public announcements, and private suspicions. In Gilboa, I., editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, pages 43–56.

[Baltag et al., 2018] A. Baltag, R. Boddy and S. Smets (2018). Group knowledge in interrogative epistemology. Springer series 'outstanding contributions to logic', volume dedicated to J. Hintikka, Springer.

[Baltag and Smets, 2008] Baltag, A. and Smets, S. (2008). A Qualitative Theory of Dynamic Interactive Belief Revision. Texts in Logic and Games, Amsterdam University Press, vol 3, pp.9–58.

[Baltag, et al., 2013] Baltag, A., Christoff, Z., Hansen, J.U. and S. Smets (2013). Logical Models of Informational Cascades. In J. van Benthem and F. Liu eds., Studies in Logic. College Publications, Vol.47, pp. 405–432.

[Baltag, et al., 2018] Baltag, A., Christoff, Z., Rendsvig, R., and S. Smets (2018). Dynamic Epistemic Logics of Diffusion and Prediction in Social Networks. *Studia Logica*, pp. 1–43, online first.

[Baltag, 2016] Baltag, A. (2016) To Know is to Know the Value of a Variable *Advances in Modal Logic*, Vol. 11, CSLI Publications. pp. 135–155.

[Belnap et al., 2001] Belnap, N., Perloff, M., and Xu, M. (2001). *Facing the Future*. Oxford: Oxford University Press.

[Bergstra et al., 2001] Bergstra, J., Ponse, A., and Smolka, S. (2001). *Handbook of Process Algebra*. Elsevier Science B.V.

[Beth, 1971] Beth, E. W. (1971). *Aspects of Modern Logic*. D. Reidel Publishing Company/ Dordecht-Holland.

[Bikhchandani et al., 1992] Bikhchandani, S., Hirshleifer, D. and Welchm, I. (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), pp. 992–1026.

[Bod, 2008] Bod,R. (2008) The Data-Oriented Parsing Approach: Theory and Application PP-2008-24, ILLC publicatons.

[Bonnema, et al., 1999] Bonnema, R., Buying, P. and Scha, R. (1999). A New Probability Model for Data Oriented Parsing in P. Dekker and G. Kerdiles ed., *Proceedings of the 12th Amsterdam Colloquium.*

[Bonizzoni et al., 2013] Bonizzoni,P., V. Brattka, B. Löwe (eds). *The Nature of Computation: Logic, Algorithms, Applications*, Proceedings of the 9th Conference on Computability in Europe (CiE 2013), LNCS 7921 Springer: Heidelberg.

[Brandenburger, 2014] Brandenburger, A. (2014). *The Language of Game Theory, Putting Epistemics into the Mathematics of Games.* World Scientific Series in Economic Theory.

[Brafman et al., 1997] Brafman, R. I., Latombe, J., Moses, Y., and Shoham, Y. (1997). Applications of a logic of knowledge to motion planning under uncertainty. *J. ACM*, 44(5):633–668.

[Bratman,1987] Bratman, M. (1987). *Intention, Plans, and Practical Reason.* CSLI publications.

[Cardon and Itmi, 2016] Cardon, A. and Itmi, M. (2016). *New Autonomous Systems, Volume 1.* John Wiley & Sons, Inc.

[Chakraborty et al., 2011] Chakraborty, P., Saxena, P. C., and Katti, C. P. (2011). Fifty years of automata simulation: A review. *ACM Inroads*, 2(4):59–70.

[Church, 1936] Church, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58 (2):345–363.

[Ciardelli and Roelofsen, 2011] Ciardelli, I. and Roelofsen, F. (2011). A knowledge based semantics of messages. *Journal of Philosophical Logic*, 40:55—-94.

[Christoff, et al., 2016] Christoff, Z., Hansen, J.U, and Proietti, C. (2016). Reflecting on social influence in networks. *Journal of Logic, Language and Information*, 25(3-4), pp. 299–333.

[Demey, et al., 2017] Demey, L., Kooi, B., and Sack, J. (2017). Logic and Probability. The Stanford Encyclopedia of Philosophy.

[Fagin et al., 1995] Fagin, R., Halpern, J., Moses, Y., and Vardi, M. (1995). *Reasoning about Knowledge*. Cambridge, MA: The MIT Press.

[French, 1956] French, JRP Jr. (1956). A formal theory of social power. *Psychological Review*, 63(3):181–194.

[Friedkin, 1998] Friedkin,N. E.(1998). *A Structural Theory of Social Influence*. Cambridge University Press.

[Gabbay, 1998] Gabbay, D. M (1998). *Fibring Logics*. Oxford University Press.

[Gabbay and Guenthner,1983-] Gabbay, D. M. and F. Guenthner (1983-). *Handbook of Philosophical Logic*. Elesvier.

[Gabbay, et al.,1993] Gabbay, D. M., C. J. Hogger and J. A. Robinson (1993-1998). *Handbook of logic in Artificial Intelligence and Logic Programming*. Oxford University Press.

[Gattinger and van Eijck, 2015] Gattinger, M. and van Eijck, J. (2015). Towards Model Checking Cryptographic Protocols with Dynamic Epistemic Logic. *Proceedings LAMAS*.

[Gödel, 1931] Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. In Feferman, S., editor, *Kurt Gödel Collected works*, pages 144–195. Oxford University Press, 1986.

[Halpern, 2005] Halpern, J.Y., (2005). *Reasoning about Uncertainty*. The MIT Press.

[Halpern and Pass, 2009] Halpern, J.Y., and Pass, R. (2009) Iterated regret minimization: a new solution concept. *Proceeding Proceedings of the 21st international jont conference on Artifical intelligence* (IJCAI'09). pp. 153–158.

[Halpern and Rego, 2009] Halpern, J.Y., and Rego, L. (2009) Reasoning About Knowledge of Unawareness Revisited. arXiv:0906.4321 [cs.AI]

[Halpern, 2016] Halpern, J.Y. (2016) *Actual Causality*. The MIT Press

[Harel, 1987] Harel, D. (1987) *Algorithmics: The Spirit of Computing*. Addison-Wesley, Reading, MA, 1987.

[Harrison, et al., 2018] Harrison-Trainor, M, Holliday, W. and Icard,T. (2018) Inferring Probability Comparisons. *Mathematical Social Sciences*.

[Harrenstein, et al, 2001] Harrenstein, P., van der Hoek, W., Meyer,J.J., Witteveen,C.(2001) Boolean games. *Proceeding of the 8th conference on Theoretical Aspects of Rationality and Knowledge*, pp. 287–298.

[Haugeland, 1997] Haugeland, J. (1997). *Mind Design II.* The MIT Press.

[Hodges, 2018] Hodges, W. (2018). Logic and Games. The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.).

[Hornischer, 2019] Hornischer, L. (2019). Trajectory domains: analyzing the behavior of transition systems. ILLC pre-print paper.

[Horty, 2014] John F. Horty (2014) *Reasons as Defaults.* Oxford University Press.

[Hintikka, 1973] Hintikka, J. (1973). *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic.* Oxford: Clarendon Press.

[Ibeling and Icard, 2018] Ibeling, D., Icard, T. (2018) On the Conditional Logic of Simulation Models. *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (IJCAI).

[Isaac et al, 2014] Isaac, A., Szymanik, J. and Verbrugge, R. (2014). Logic and Complexity in Cognitive Science In Baltag, A. and Smets, S. (eds). *Johan van Benthem on Logic and Information Dynamics*, p.787–824, Springer.

[Kelleher, et al., 2015] Kelleher J.D., Namee,B.M., D'Arcy.D. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies.* The MIT Press.

[Kelly, 1996] Kelly, K. (1996). *The Logic of Reliable Inquiry.* Oxford University Press, USA.

[Klein and Rendsvig, 2017] Klein, D., Rendsvig, R.K. (2017). Convergence, Continuity and Recurrence in Dynamic Epistemic Logic In A. Baltag, J. Seligman and T. Yamada (eds.), *Logic, Rationality, and Interaction* (LORI 2017). Springer. pp. 108-122 (2017).

[Kneale and Kneale, 1962] Kneale, W. and Kneale, M. (1962) The Development of Logic. Oxford University Press: New York.

[Kremer and Mints, 2005] Kremer, P. and Mints, G. (2005) Dynamic topological logic. *Annals of Pure and Applied Logic* 131 (1-3):133-158.

[Leitgeb, 2004] Leitgeb, H. (2004). Inference on the Low Level: An Investigation into Deduction, Nonmonotonic Reasoning, and the Philosophy of Cognition. Kluwer Academic Publishers.

[Leitgeb, 2017] Leitgeb, H. (2017). The Stability of Belief, How Rational Belief Coheres with Probability. Oxford University Press.

[Liang and Seligman, 2011] Liang, Z. and Seligman, J (2011). The dynamics of peer pressure. *Proceedings, volume 6953 of Lecture Notes in Computer Science*, pp. 390–391. Springer.

[Lorini, 2018] Lorini, E. (2018) In Praise of Belief Bases: Doing Epistemic Logic Without Possible Worlds. in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp.1915–1922.

[Liu, 2011] Liu, F. (2011). *Reasoning about Preference Dynamics*, volume 354 of *Synthese Library*. Springer.

[Liu, et al., 2014] Liu, F., Seligman, J. and Girard, P. Logical dynamics of belief change in the community. *Synthese*, 191(11), pp. 2403–2431.

[Maynard Smith, 1982] Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press.

[Minsky, 1961] Minsky, M. (1961). Steps toward artificial intelligence. In *Proceedings of the IRE*, volume 49, pp. 8–30.

[Nisan et al., 2007] Nisan, N., Roughgarden, T., Tardos, E., and Vazirani., V. V. (2007). *Algorithmic Game Theory*. Cambridge University Press, New York, NY.

[Nishimori, 2001] Nishimori, Hidetoshi (2001). Statistical Physics of Spin Glasses and Information Processing: An Introduction. Oxford: Oxford University Press.

[Papadimitriou, 1994] Papadimitriou, C. (1994). *Computational Complexity*. Addison Wesley.

[Parikh and Ramanujam, 2003] Parikh, R. and Ramanujam, R. (2003). A knowledge based semantics of messages. *Journal of Logic, Language and Information*, 12:453–467.

[Perea, 2019] Perea, A. (2019). Epistemic game theory. In the *Handbook of Rationality*. To appear.

[Premack and Woodruff, 1978] Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, 4, pp. 515–526.

[Rao and Georgeff, 1991] Rao, A. and Georgeff, M. (1991). Modeling rational agents within a BDI-architecture. In Allen, J., Fikes, R., and Sandewall, E., editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pp. 473–484. San Mateo, CA: Morgan Kaufmann.

[Osborne and Rubinstein, 1994] Osborne, M. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT Press.

[Pearl, 2000] Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

[Popper, 1963] Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge, London.

[Restall, 2000] Restall, G., (2000). *An Introduction to Substructural Logics*. London: Routledge.

[Restall, 2005] Beall,J.C. and Restall,G. (2005). *Logical Pluralism*. Oxford University Press.

[Russel, et al., 1994] Norvig, P., S., Russell (1994) Artificial Intelligence: A Modern Approach, 3rd ed. Prentice Hall

[Seligman et al., 2011] Seligman, J., Liu, F. and Girard P (2011). Logic in the community. *Proceedings of ICLA 2011*, vol. 6521 of LNCS, pp. 178–188. Springer.

[Seligman et al., 2013] Seligman, J., Liu, F. and Girard P (2013). Facebook and the Epistemic Logic of Friendship *Proceeding of TARK*, arXiv:1310.6440 [cs.LO]

[Shi, 2018] C. Shi (2018). *Reasons to Believe*. Ph.d Dissertation, ILLC, University of Amsterdam, 2018.

[Shoham and Leyton-Brown, 2008] Shoham, Y. and Leyton-Brown, K. (2008). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations* Cambridge University Press

[Skyrms, 2010] Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press, USA.

[Smets and Velazquez-Quesada, 2017] Smets, S. and Velázquez-Quesada, F.R. (2017). How to make friends: A logical approach to social group creation. *Proceedings of LORI 2017*, volume 10455 of LNCS, pp. 377–390. Springer.

[Smullyan, 1994] Smullyan, R. (1994). *Diagonalization and Self-Reference*. Oxford University Press.

[Spohn, 1988] Spohn, W. (1988). Ordinal conditional functions: a dynamic theory of epistemic states. In Harper, W. and Skyrms, B. eds., Causation in Decision, Belief Change, and Statistics, vol. II, pp.105-–134.

[Spohn, 2012] Spohn, W. (2012). The Laws of Belief: Ranking Theory and Its Philosophical Applications. Oxford University Press.

[Street, 2005] Street, T. (2005). *Avicenna*. The Islamic Texts Society, Cambridge, UK.

[Swayamdipta et al., 2018] Swayamdipta, S., Thomson, S., Lee, K., Zettlemoyer, L., Dyer, C., and Smith, N. A. (2018). Syntactic scaffolds for semantic structures. arXiv:1808.10485 [cs.CL].

[Talbott, 2016] Talbott, W. (2016). Bayesian Epistemology. The Stanford Encyclopedia of Philosophy.

[Turing, 1936] Turing, A. (1936). On computable numbers, with an application to the entscheidungsproblem. In *Proceedings of the London Mathematical Society*, vol.42, pp. 230–265.

[Turing, 1950] Turing, A. (1950). Computing machinery and intelligence. *Mind*, 236:433–460.

[Väänänen, 2007] Väänänen, J. (2007). *Dependence Logic A New Approach to Independence Friendly Logic* Cambridge University Press

[van Benthem, 2009] van Benthem, J., Gerbrandy, J. and Kooi, B.(2009). Dynamic Update with Probabilities. *Studia Logica*, 93: 67–96.

[van Benthem, 2011] van Benthem, J. (2011). *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge.

[van Benthem, 2014] van Benthem, J. (2014). *Logic in Games*. The MIT Press.

[van Benthem, 2015] van Benthem, J. (2015). Oscillations, Logic, and Dynamical Systems ILLC Publications: PP-2015-10

[van Benthem and Smets, 2015] van Benthem,J., S. Smets. (2015). Dynamic Logics of Belief Change. In H. van Ditmarsch, J.Y. Halpern, W. van der Hoek and B. Kooi (Eds.). *Handbook of Logics for Knowledge and Belief*, College Publications, pp.313-393.

[van Benthem, 2018] van Benthem, J. (2018). Computation as social agency: What, how and who (2018) *Information and Computation*, Vol. 261, Part 3, pp. 519–535.

[van Benthem, 2018] Implicit and Explicit Stances in Logic (2018). *Journal of Philosophical Logic*, pp. 1–31.

[van Benthem et al., 2015] van Benthem, J., Ghosh, S., and Verbrugge R. Eds (2015). Models of Strategic Reasoning: Logics, Games, and Communities. FoLLI series, Volume 8972 of LNCS, Springer: Berlin.

[van Benthem and Klein, 2019] van Benthem, J., D. Klein. Logics for Analyzing Games *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition)*, Edward N. Zalta (ed.).

[van Benthem and Pacuit, 2011] van Benthem, J. and Pacuit, E. (2011). Dynamic logics of evidence-based beliefs. *Studia Logica*, 99(1-3):61–92.

[van Benthem, et al., 2018] van Benthem, J. and van Eijck, J. and Gattinger, M. and Su, K. (2018). Symbolic Model Checking for Dynamic Epistemic Logic – S5 and Beyond. *Journal of Logic and Computation*, 28(2):367–402.

[van Ditmarsch, 2019] van Ditmarsch, H. (2019). Doxastic and epistemic logic. In the *Handbook of Rationality*. To appear.

[van Ditmarsch et al., 2007] van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*. Springer: Berlin.

[van Harmelen et al., 2008] van Harmelen, F., Lifschitz, V., Porter, B. W. (2008). *Handbook of Knowledge Representation*. in *Foundations of Artificial Intelligence*. Elsevier, 2008.

[van Lambalgen, 1996] van Lambalgen, M. (1996). Randomness and foundations of probability: von Mises' axiomatisation of random sequences. in *Statistics, probability and game theory*, pp. 347–367, Institute of Mathematical Statistics, Hayward, CA.

[van Rooij and Schulz, 2019] van Rooij, R. and Schulz, K. (2019). Conditionals, Causality and Conditional Probability. *Journal of Logic, Language and Information* 28 (1):55-71

[Sandu, 2012] Sandu, G. (2012). Independendly-Friendly Logic: Dependence and Independence of Quantifiers in Logic *Philosophy Compass* 7 (10):691-711

[Su and Sattar, 2008] Su, K. and Sattar, A. (2008). An Extended Interpreted System Model for Epistemic Logics. in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*. 554–559.

[Wang, 1996] Wang, H. (1996). *A Logical Journey: From Gödel to Philosophy*. The MIT Press, Cambridge MA.

[Wooldridge, 2009] Wooldridge, M. (2009). *An Introduction to MultiAgent Systems*. John Wiley & Sons, Inc.

[Wooldridge, 2010] Wooldridge, M. (2010). *Reasoning about Rational Agents*. The MIT Press, Cambridge MA.

[Xue, 2017] Xue, Y. (2017). *In Search of Homo Sociologicus*. Ph.D dissertation, The Graduate Center, City University of New York, 2017.

[Yamada, 2008] Yamada, T. (2008). Logical dynamics of some speech acts that affect obligations and preferences. *Synthese*, 165(2):295–315.