The Study of Informational Processes

an introduction to logic

Smets, S.; Velázquez Quesada, F.R.

DOI
10.26116/secondthoughts-sie-engelen-openpresstiu-2021-smets-velazques-quesada-1
Publication date
2021
Document Version
Final published version
Published in
Second Thoughts
License
CC BY-NC-ND

Second Thoughts

First Introductions to Philosophy

Edited by
Mareen Sie & Bart Engelen

Web Version

Open Press TiU

DOI 10.26116/second-thoughts-openpresstiu-jote

Tilburg University Press,
Tilburg, The Netherlands
https://www.openpresstiu.org

Design and Typesetting by Thomas F. K. Jorna, JOTE Publishers

Contents

Sonja Smets & Fernando R. Velázquez Quesada

The Study of Informational Processes

An Introduction to Logic

1 • Introduction

Suppose your teacher tells you 'if you do study hard, then you pass the exam'. If you consider the teacher's statement to be true and conclude that 'if you do not pass the exam, then you do not study hard', you have made a valid inference. However classical logic will tell you that it wouldn't be a valid inference if you concluded instead that 'if you pass the exam, then you do study hard'. Indeed, logic is typically understood as the study of *inference*, the process of deriving a conclusion from a set of premises. Even more: for some, logic is actually only about *valid* inferences: those in which the truth of the premises guarantee the truth of the conclusion. The study of valid inference has been one of the main focuses within logic, and in fact it may have been its original motivation. But, through its contact with other domains in philosophy and with different fields, such as economics, computer science and artificial intelligence, logic has become so much more. Modern logic can be understood as the study of *(i)* the different information-changing *actions* that real and/or artificial *agents* (e.g., human beings, but also computer programs) can perform, *(ii)* the way in which these actions *affect* the agents' individual and collective *information*, and *(iii)* the way in which these actions *interact* with one another, creating long-term information-changing procedures. The aim of this chapter is to provide a brief storyline describing how logic has evolved from focusing on individual human reasoning patterns, such as deciding whether proposition "*q*" follows from both "*p*" and "*p* implies *q*", to analyzing the flow of information in complex social scenarios where human and artificial agents interact.

2 • PATTERNS OF REASONING

Aristotle (384-322 BC) is considered to be the father of Western logic, his theory of syllogisms has had an unparalleled influence on the history of Western thought. In this theory, Aristotle focused on what he called *deduction*, a concept that nowadays is also called valid inference: a form of inference in which the conclusion follows *necessarily* from the premises.[1] His goal was not to study the properties of inferences used by regular people (what one might call a *descriptive* account); it was rather to identify inferences that people *should* use (what one might call a *normative* account).

One of the features that made the theory of syllogisms so significant is that, when identifying valid inferences, Aristotle focused not on the actual content of the involved statements; rather, he focused on the inference's *form*. For example, when considering an inference such as "Every Greek is human; every human is mortal. Therefore, every Greek is mortal", Aristotle did not discuss the meaning of the terms "Greek", "human", and "mortal". He rather directed his attention towards the more general form "Every X is Y; every Y is Z. Therefore, every X is Z", then observing that *any* inference following this scheme is valid, regardless of what X, Y and Z are. This was a simple but powerful idea, as it allowed his findings to be suitable for all areas of inquiry. Indeed, Aristotle's framework "... systematizes the principles licensing acceptable inference, and helps to highlight at an abstract level seductive patterns of incorrect inference to be avoided by anyone with a primary interest in truth" (Shields, 2016).

Aristotle's study of syllogisms was devoted to inferences that have two *premises*, three *properties* (the X, Y and Z of before), and in which premises and conclusion are all statements of the form "(it is not the case that) every/some X is (not) Y". Yet, the crucial insight of looking for valid *patterns* of inference is probably its most important legacy. From there, several lines of work followed, extending the idea by studying patterns that are not restricted to the 'simple' forms of reasoning a syllogism allows. For example, one might reason from more than two premises. Similarly, the

[1] More precisely, a form of inference in which the conclusion is true *in every situation* in which all the premises are true. Equivalently, a form of inference in which *it is not possible* for all the premises to be true and yet for the conclusion to be false. Note thus how, in a valid inference, the premises do not need to be true: it is only required that, *if* all the premises are true, then the conclusion *must* be true.

premises and the conclusion might involve more than three properties (or maybe less). More interestingly, one might use *connectives*, such as *conjunctions* ("and"), *disjunctions* ("or") and *negations* ("not"), to create more complex statements. Also interesting would be to require a more complex *quantification* pattern than the one present in syllogisms. For this, note how every statement in a syllogism uses exactly one *quantification* over the objects in the domain of discourse, being either a *universal quantifier* ("every") or an *existential quantifier* ("some"). However, certain areas of inquiry may require more complex combinations, as in case when one wonders whether "somebody is loved by everybody" guarantees that "everybody loves somebody".

Among the several directions in which the *language* used in syllogisms can be extended, two of them are arguably the most important. The first one, called nowadays *propositional logic*, focuses on the valid reasoning patterns that arise when the premises and the conclusion are built from basic *propositions* (statements that can be qualified as true or false) by means of *connectives* (the already mentioned conjunction, disjunction and negation, and also others, such as implication). For example, within propositional logic one can prove that the famous *modus ponens* (if both "*p*" and "*p* implies *q*" hold, then so does "*q*") is a valid form of inference. The second one, called nowadays *predicate logic*, focuses on the valid reasoning patterns that arise when the premises and the conclusion are built from basic *predicates* (statements whose truth-value depend on the properties of and relations between the *objects* in the domain of discourse) by means of the mentioned connectives and also arbitrary patterns of universal and existential *quantification* over objects.[2] For example, within predicate logic one can prove that the following inference is valid: if "somebody is loved by everybody", then surely "everybody loves somebody". Sharing the use and interpretation of connectives (also called *Boolean connectives*), both propositional logic and predicate logic coincide in several aspects.[3]

[2] Predicate logic's full name is *first-order* predicate logic, emphasizing the fact that, while quantifiers can act over objects ("all objects have the property *P*"), they cannot act over *sets* of objects ("all sets of objects have the property *P*"). Quantification over sets of objects is studied in *second-order* predicate logic and *higher-order* predicate logic (the latter includes quantification over objects, over sets of objects, over sets of sets of objects, and so on).

[3] In fact, predicate logic can be seen as an *extension* of propositional logic.

Together, they constitute the core of what is typically called *classical logic*.

Even though classical logic is based on a number of strong principles, these principles may not always seem so natural when applied to different reasoning scenarios. Hence, some of the essential assumptions and principles of classical logic can be challenged; this gives rise to a variety of systems commonly referred to as *non-classical logics*. For example, classical logics rely on the idea that every sentence is either true or false,[4] but one can allow them to have other truth-values.[5] A further departure from classical logic encompasses the systems falling under the umbrella of *modal logic,* in which we allow different modes of truth such as necessarily true or possibly true. These systems are extensions of propositional logic that use additional connectives called *modal* operators (e.g., the *unary* operator "$\Box$"). Depending on their particular interpretation, expressions of the form "$\Box p$" can be read, e.g., as "p is necessary" (*alethic* modal logic), "p is obligatory" (*deontic* logic) or "from now on, p will always be the case" (*temporal* logic). In the context of an agent and its information, two crucial readings are "p is believed" (*doxastic* logic) and "p is known" (*epistemic* logic). One of the reasons for the importance of (basic) modal logics is that they can be seen as *classical* logics located between propositional logic and predicate logic.

The identification of valid patterns of reasoning has been a fundamental tool in sciences with a primary interest in filling in the 'logical consequences' of what we already know. Hence the importance of logic in, e.g., mathematics and computer science. Nevertheless, the view that logic should be only about one notion of inference (let alone *valid* inference) is rather restrictive and is not in line with the work of the great pioneer Bernard Bolzano (1781-1848), who thought that logic should rather chart the many different consequence relations that people might use, depending on the reasoning task at hand. A similar rich view is found in the work of other philosophers, as John Stuart Mill (1806-1873) and, especially,

[4] Sentences of the form "p or not p" and "if both p and not p hold, then q holds" are always true in classical logic, regardless of what p and q represent. Yet, these principles are dropped, respectively, in *intuitionistic logic* and in *paraconsistent logic*.

[5] The most popular of such logics include not only the systems in which sentences can have one of *three* possible truth-values (*three-valued logics*), as initially developed by the logician Jan Łukasiewicz (1878-1956), but also those in which the truth-value of a sentence can be any real number between 0 and 1 (*fuzzy logic*).

C. S. Peirce (1839-1914). This interest in a wider variety of inference patterns was encouraged in the early 1980s, mainly through the emergence of artificial intelligence and its search for reasoning patterns that are similar to how humans solve problems. Inferences such as "the grass is wet; therefore, it rained" (*abductive reasoning*, a form of inference that looks for an explanation of an unpredicted observation) and "Pingu is a bird; therefore, it flies" (*default reasoning*, a form of inference usually used in situations with incomplete information) have become the subject of study in modern logic. Such studies are linked to investigations in the empirical sciences where we look for explanations or reason about what is typically (but not always) the case. Note how both inferences are non-valid in classical logic: one can easily think of a situation in which the grass is wet and yet it did not rain (e.g., "the sprinkler is on"), and also of a situation in which Pingu is a bird and yet it does not fly (e.g., "Pingu is a penguin"). In other words, both forms of inference are non-valid because the conclusion might be dropped in the light of further information.

By looking at forms of reasoning whose conclusions might be dropped when further information arrives (*non-monotonic* reasoning), logic has also expanded its original goal. Its focus is no longer only normative, looking for the reasoning patterns people *should* use. Logicians nowadays are also interested in representing and understanding the many other 'fallible' forms of inference that 'real' non-omniscient humans use every day. Note that this does not undermine the search for valid patterns; it rather *enriches it*. In a classical experiment, the Wason selection task, subjects are presented with four cards placed on a table, with each card having a letter on one side and a number on the other. The visible faces of the cards show A, K, 4 and 7, and the subjects are asked to point to the card(s) that must be turned over to test whether the rule "if there is a vowel on one side, there is an even number on the other" holds. Most subjects point (correctly) to the card showing A (if the rule holds, its other side should have an even number). However, while most subjects also point (incorrectly) to 4 (the rule does not enforce anything when a side has an even number), only few of them point (correctly) to 7 (if the rule holds, its other side should *not* have a vowel). These results are surprising and have given rise to a series of replication studies and variations of the selection task. Yet, the conclusion from these studies should not be that humans fail to reason 'correctly' or that real-life reasoning is very different from what our theories of logic say.

First, when subjects are explained the logical solution of the Wason selection task, they understand it and agree with it. Second, subjects presented with versions of the task based on 'more familiar' rules (e.g., "if you drink alcohol here, you have to be over 18") tend to point to the correct cards. In order to explain this, it has been suggested that the *representation* of the reasoning task plays an important role, and one can also look for the governing principles.[6] George Boole (1815-1864), often considered the father of the purely mathematical approach to propositional logic, regarded the rules of propositional calculus as describing essential human thought. Nevertheless, he acknowledged that human reasoning often deviates from this canon, taking this as a sign that there are further laws that still need to be discovered.

Recognising the important role played by non-monotonic reasoning allows logic to look at 'more real' agents. Still, our story so far only focuses on the inferences a single agent can obtain on its own. As such, it highlights a single person's intellectual effort, as paradigmatically exemplified by Auguste Rodin's sculpture *The Thinker*. However, as the next section explains, individual inferences are not the only meaningful informational action.

3 • REASONING IN INTERACTION

Consider the following situation:

> "You are in a restaurant with your parents, and you have ordered fish, meat, and vegetarian, for you, your father and your mother, respectively. A new waiter comes with the three dishes. What can he do to know which dish corresponds to which person?"

The waiter can ask "Who has the fish?"; then he can ask "Who has the meat?". Now he does not have to ask anymore: as the logician Johan van Benthem [b.1949] points out, "two questions plus one inference are all that is needed".

The example shows how the information flows when agents start *interacting* with their environment. In the example, the answers to the

[6] One explanation that has been offered is that this scenario points to the presence of a classical framing effect, a cognitive bias studied in psychology.

posed questions provide the waiter with information that he otherwise would have not been able to obtain. Maybe more importantly, he can make use of the information he has received to perform further inferences that, together with his initial information (each person gets only one dish, and each dish corresponds to one person), allows him to deduce who gets the vegetarian dish. In other words, it is the *interplay* of these processes, the 'external' asking-questions-then-receiving-answers and the 'internal' inference, that allows him to get to know what the real situation is.

Despite the fact that logic has looked into informational acts other than inference only relatively recently, the realization that there are further informational actions that should be taken into account was already present in earlier days. In fact, it was already discussed by Aristotle. At the time, it was argued that knowledge could only arise from valid reasoning based on known premises. But if that is true, the only way in which the premises of a valid reasoning can be known is if each one of them is, in turn, the conclusion of some other valid inference. This argument can be used again, now on the newly found inferences, and its repetition becomes then a process that tries to find the 'original' premises from which everything else follows. But then, there is a problem. On the one hand, if the process never ends, then nothing can be demonstrated. On the other hand, if the process ends, then these 'original' premises are not properly justified, and consequently neither are any of the statements deduced from them. For solving this *regress problem*, Aristotle argues [*Posterior Analytics* I.2, II.19] that the idea that scientific knowledge is only possible by using valid reasoning with scientifically known premises is incorrect: there is another form of knowledge for 'the original' premises, and this provides the starting point for demonstrations. Aristotle compares the capacity of having this knowledge with the capacity for sense-perception, which can be understood as the capacity of receiving information from *external* sources. In this sense, it can then be said that information (or, in the central terms of epistemology, knowledge and beliefs) arise not only from inferences, but also from the interaction with our environment.

With this idea in mind, one can provide a more precise description of the way the information of the waiter (his information state) changes through his interaction with the family. One possible way of describing this process, useful later, is the following. At the start, the new waiter has full uncertainty about who ordered what, yet he knows that three people each ordered a different dish (fish, meat or vegetarian). This leaves open six possible ways in which the world could be (all possible ways in which three dishes can be distributed over three people), with only one of them having the correct person-dish distribution. Without any further information, our waiter cannot distinguish between these six 'possible worlds'. It is only after the first question-answer, when our waiter learns who ordered fish, that he can discard some of these alternatives (in fact, four of them). When he receives the answer to his second question, he can discard one of the two remaining possibilities, thus gaining full knowledge of the correct person-dish allocation.

The waiter thus gets his information from external sources in the form of answers to his questions. As we saw, Aristotle acknowledged the importance of external sources by referring to sense perception. Use of the senses, however, is a limited form of interaction. Sense-perception indeed allows us to get information from external sources, but these external sources are then being understood as 'unresponsive' objects. The informational acts that a human can carry out with these objects can be called *observations*; as useful as they are, they are not the only way in which one can interact with the environment.

In the restaurant example, the new waiter does not have an interaction with objects. He has an interaction with other agents, who might indeed provide him with further information, but who (just as himself) have also knowledge, beliefs, preferences and goals, and therefore also *react* and *change*. In these cases, this exchange can now be properly called *interaction*, and it goes beyond what has been described above. When the waiter asks a question, he reveals the family not only that he does not know the answer, but also that he wants to know the answer, and that he considers it possible for the family to know it. Then, in turn, the family answers because they want to start their dinner, and they know their answer will allow the waiter to know who gets each dish.

It is important to notice that, during the conversation, the new waiter is not the only one whose information changes: the information of everyone else present also changes. If the answers given in reply to the waiter's questions were truthful and truly public (everybody witnesses the action, everybody sees everybody witnessing the action, and so on), then the person-dish distribution has become *common knowledge* among everybody present: everybody knows who gets each dish (in particular, the waiter received the needed information), everybody knows that everybody knows (in particular, the answers of the family were public, so they know the waiter knows), everybody knows that everybody knows that everybody knows, and so on.[7] Thus, they all get to know something new; in particular, they all get *higher-order* knowledge, that is, knowledge about the knowledge of the agents involved (including themselves). It is worthwhile to notice how common knowledge would not have been reached if the waiter's questions were answered only by one member of the family while the others were not looking or listening: the inattentive agents would not know that the waiter knows.

It is also worthwhile to notice that, while in this scenario the interaction is successful (the important bits of information become common knowledge), the outcome could have easily been otherwise if the agents behaved slightly differently. For instance, if the waiter had other goals, he might have attempted to mislead by asking insincere questions. Likewise, if the members of the family wanted to confuse the waiter, their answer could have been untruthful, or they could have not answered at all. These possibilities are what distinguish 'simple' acts of observation, with one side being an unresponsive object (as in sense perception), from real acts of interaction, with both sides being agents with knowledge, beliefs, preferences and goals.

4 • MODELING INFORMATION AND INFORMATION CHANGE

If the main focus of logic is the way different information-changing actions affect the information of agents, then its formal study requires at least two main ingredients. We need a formal representation of the information that the agents have, and we need a formal representation of the effects of these actions. For the first, the above-mentioned *epistemic logic* (see, e.g., Fagin et al., 1995; van Ditmarsch et al., 2015) is an important tool, as it allows us to represent and reason about the knowledge (or, when understood in a broader sense, the different kinds of knowledge and the different forms of belief) of both individual agents and groups of

[7] Public actions are the paradigmatic way of reaching common knowledge. For example, when cash is used to pay for a product or a service, both the client and the cashier tend to put the money (the payment and the change, respectively) in a tray. In this way, both have common knowledge about the amount of money being exchanged. Similarly, when applying an injection, the specialist will unwrap the syringe and needle in public view, so it will be common knowledge that they are brand new.

them. For the second, the ideas behind *dynamic epistemic logic* (see, e.g., van Benthem, 2011) are insightful, as they allow us to describe changes in the structures representing the agents' information. In fact, the basics of both epistemic logic and dynamic epistemic logic have been already described, albeit implicitly, when analysing the restaurant example. Here it is a more precise presentation of these ideas.

To represent the knowledge of an agent, we adopt the basic idea of *information as range*. An agent's available information is essentially represented as a range (collection) containing all the different ways the world could be from her perspective (i.e., all worlds the agent considers possible). In settings that follow this idea, the *knowledge* of an agent is defined as what holds in all those alternatives, that is,

> an agent knows that a certain proposition is true whenever the proposition is true in *all* worlds/situations in her range.

Note how, in this approach, the model does not represent directly the knowledge of the agent: it rather represents the agent's *uncertainty*, and then defines her knowledge in terms of it. In the restaurant example, the initial information of the new waiter can be modelled by a range containing the six possible ways in which the three dishes can be distributed over the three persons.

The discussion above suggests assigning a set of 'possible worlds' to each agent. Yet, it is more practical to represent the range by means of a relation connecting worlds to worlds. Thus, given a possible world, the agent's *indistinguishability* relation indicates which are the worlds that the agent considers possible from it. These structures for representing an agent's knowledge, called *possible worlds model*, *Kripke models* (after the philosopher Saul Kripke [b.1940]) or, more generically, *relational models*, have several advantages. An important one is that the move from a single-agent case to a multi-agent scenario becomes straightforward: we only need a separated depiction of the worlds each agent considers possible, and this can be done by means of different indistinguishability relations. A more interesting advantage is that, thanks to the relational representation, these structures describe not only what an agent knows about the real world, but also what they know about what *they and other agents* know about the real world, and so on. As discussed above, this form of *higher-order* knowledge is crucial when an agent interacts with others.

It is important to notice that the notion of knowledge depicted by these structures is very strong. Before considering an example, it is worthwhile to briefly mention some of the features of this form of knowledge. The most noticeable characteristic is that knowledge is 'closed' under modus ponens. Indeed, if an agent knows both "*p*" and "*p* implies *q*", then both are true in every world in the agent's range; but then, "*q*" is also true in each one of those worlds, and hence it is known by the agent. Thus, agents whose knowledge is represented in this way do not need to perform modus ponens: they already have everything they would get from it.[8] Another important feature of relational models refers to the consequence of making the assumption that every world is indistinguishable from itself, which implies that the agent always includes the real world in her range. Under this natural and seemingly harmless postulate, whatever is known by the agent *needs to hold in the real world*, and therefore cannot be revised (this is called among philosophers *irrevocable* knowledge). Finally, under other reasonable assumptions about the indistinguishability relation (and hence the range), the knowledge of an agent is both *positively and negatively introspective*. This means that, if the agent knows that something is true, she knows that she knows it (positive introspection) and if she does not know that something is true, she knows that she does not know it (negative introspection). All of these features can and have been debated extensively, with epistemologists proposing additional fine-grained epistemic attitudes (e.g., explicit knowledge, defeasible and/or non-introspective knowledge, different forms of beliefs) and logicians proposing different structures to represent them (e.g., plausibility models, neighbourhood models, evidence models). Yet, the strong form of knowledge represented by relational models is enough to exemplify our main point: agents have knowledge, and their knowledge can change when they communicate with one another.

Since an agent's knowledge is defined by its range, actions that change her knowledge can be represented by operations that change her range. In particular, consider the action that provides the agents with truthful information via public communication. By stating that a given proposition

[8] This also explains the mismatch between the initial analysis of the restaurant example (which indicated that two questions plus one deductive inference were needed) and the more detailed description provided later (in which the waiter only needed the answers to his two questions).

is true, the action is effectively ruling out all those situations (in this case, possible worlds) in which the proposition is false. Thus, an act of public observation can be represented by an operation that *reduces* the agents' range. Note how this works as one intuitively expects: by eliminating worlds, the operation decreases the agent's uncertainty, thus increasing her knowledge. In the restaurant example, this is exactly the operation used for representing the way the answers of the family affect the new waiter's information state.

Here, a word of caution is needed. A range-reducing operation is adequate for representing acts of *truthful* and *public* communication. But, if the communication were public and yet not truthful, then worlds could not be eliminated; they would need to be somehow 'downgraded', so they do not play a role when defining the information the agent is truly entertaining, and yet they remain available in case the provided information is proven to be false. In fact, if the communication is not truthful, one might rather stop calling the agent's information *knowledge*, and start calling it *belief*. Likewise, if the communication were truthful and yet not public, again worlds could not be eliminated, but now for a different reason. The agents receiving the information certainly should not take those worlds into account anymore when looking for what they know about the real world; yet, the agents that do not receive the information certainly cannot discard them. More importantly, even agents that received the information need those worlds: they need them to evaluate what they know *about what is known by agents that did not received the information*. But once again, despite these caveats, the range- reducing representation is enough to make our point (knowledge changes), and in fact also enough to find subtle details about the way information flows.

To explain the mechanism of knowledge updates and information change in more detail, here is another example: the high-school Math Olympiad puzzle called 'Cheryl's birthday', which was part of the Singapore and Asian Schools Math Olympiad 2015 Contests. In it, several agents are given both private and public information; in order to solve the problem, they have to reason about both the facts at hand and the information states of others. The puzzle reads as follows:

> Albert and Bernard have just become friends with Cheryl, and they want to know when her birthday is. Cheryl gives them a list of 10 possible dates:
>
> - May 15, May 16, May 19,
> - June 17, June 18,
> - July 14, July 16,
> - August 14, August 15, August 17.
>
> Cheryl then tells Albert the month of her birthday but not the day, whereas she tells Bernard the day but not the month. Afterwards, the following conversation takes place:
>
> - *Albert*: I don't know when Cheryl's birthday is, but I know Bernard does not know either.
> - *Bernard*: At first I didn't know when Cheryl's birthday is, but I know it now.
> - *Albert*: Then I also know when Cheryl's birthday is.

The question to the reader is: *When is Cheryl's birthday?* The correct answer is July 16$^{\text{th}}$. But how can we be so sure about it? In order to solve the puzzle, we have to reason about the information state of each of the agents. In other words, we need to ask: what does each agent know? However, just representing the agents' knowledge will not do: we also have to reason about the effect that the acts of communication have on their information state. In the analysis below, we make use of the above described tools of epistemic and dynamic epistemic logic.[9] To start, one first needs to analyse who knows what exactly before the conversation between Albert and Bernard takes place; this will be done using a possible worlds model, similar to the way we described the different possible worlds in the restaurant example. Cheryl has placed 10 options on the table, so we start with 10 possible worlds (labeled from *s1* to *s10*), each one of them representing a possible date for Cheryl's birthday: in *s1* her birthday is on May 15$^{\text{th}}$, in *s2* it is on May 16$^{\text{th}}$, in *s3* it is on May 19$^{\text{th}}$ and so on. With the possible worlds provided, it is possible to indicate Albert and Bernard's (different) ranges by providing their respective indistinguishability relations. For instance,

[9] One can graphically represent both Albert and Bernard uncertainty at each stage of the conversation about Cheryl's birthday. Section 1 of Appendix B in Baltag & Renne (2016) provides us with the required drawings (Figures B1-B4) that will help the reader to follow the different reasoning steps that we describe in this section.

Albert knows the month of Cheryl's birthday. Thus, he can distinguish any 'May'-world from any 'non May'-world; likewise for any other month. However, he does not know the day of Cheryl's birthday. From his perspective, all 'May'- worlds (*s1, s2* and *s3*) are indistinguishable from one another; likewise for any other month.

Let us now see what happens to our model when the above described conversation between Albert and Bernard takes place. Albert's first statement reveals his own uncertainty about Cheryl's birthday but adds a very important piece of information: he knows that Bernard does not know it either. Recall that Bernard knows the day (but not the full date) of Cheryl's birthday. Hence, if the day were 18^{th} or 19^{th}, he would immediately know the month too (and hence the full date); this is because each of these numbers is true in only one possible word, and thus Bernard would know immediately which the real world is. Thus, for Albert to know that *Bernard doesn't know*, it has to be so that, in *all* worlds that Albert considers possible, Bernard doesn't know the date.

Now, in case Albert would have been told the month was May, then he would not have known that Bernard doesn't know the date: in May, there is a possible world in which Bernard *knows* (19^{th}). The same reasoning holds for June, as it contains a day (18^{th}) in which Bernard knows the full date. But Albert does indeed know that Bernard doesn't know the exact date, so Cheryl's birthday cannot be in May, and neither in June. Thus, when Albert announces (we assume truthfully, and definitively publicly) that he knows Bernard does not know, he is effectively announcing that the birthday cannot be in May or June. Thus, both Albert and Bernard have now only five possible worlds left to reason about: the possible dates in July and August.

The conversation is continued by Bernard, who announces publicly that *now* he knows Cheryl's birthday. If this is a truthful statement, then the day cannot be the 14^{th} (as Bernard would not know the full date: he would not know whether the month is July or August). No other options can be discarded: for any of the other days that are still left, if that is the true one, then Bernard knows it. This leaves us with a model with only three possible worlds left.

In the final communication step, Albert says that *now* he also knows when Cheryl's birthday is. In the current model with three possible worlds, only in one of those possible worlds it is true that Albert knows the full date: the 16^{th} of July. Thus, under the assumed truthfulness of all involved parties, this is which Cheryl's birthday is.

Note how this analysis of the birthday scenario depends on a number of assumptions which are crucial to reach the conclusion we have drawn. For instance, again, it is implicitly assumed that the agent's information refers to their *irrevocable knowledge* (that is, knowledge that cannot be revised). It is also assumed that Cheryl, Albert and Bernard have *common knowledge* about a number of basic facts. It is common knowledge that Albert and Bernard did not know Cheryl's birthday *before* Cheryl handed them the 10 possible options; it is also common knowledge that both Albert and Bernard know that the correct answer is among the 10 given options. Further, it is also common knowledge that, immediately before his conversation with Bernard, Albert knows the precise month, but nevertheless does not have information about the day other than what is implied by the 10 listed dates. In other words, at this stage Albert knows the month, but not the day. Analogously, what Bernard knows exactly before his conversation with Albert is commonly known too. Moreover, it is assumed that Albert and Bernard tell the truth when they speak and are fully heard and understood in their public conversation. Finally, it is also assumed that Albert and Bernard have immediate access to every deductive consequence of what they know. Any deviation of these assumptions may lead to a different outcome of the problem.

Yet, the formal tools used through the analysis allow us to look at subtle details that otherwise could not have been noticed. One of them is relative to notions of knowledge for *groups* of agents. The notion of common knowledge has been already mentioned, but there are other important group epistemic attitudes. One of them is that of *distributed knowledge*: the knowledge obtained when a group of agents put together all they individually know and include all its logical consequences (i.e., everything that can be *deduced* from it). In the example, Cheryl's birthday is distributed knowledge among Albert and Bernard once Cheryl has told Albert and Bernard, separately and respectively, the month and the day. Are there other relevant group attitudes? Concepts like these are important when modelling, e.g., distributed systems within computer science.

Note also how, from his perspective, Albert and Bernard's conversation can be seen as a successful communication process: it transforms their distributed knowledge of the full date of Cheryl's birthday into common knowledge. But then, one can wonder: what if Bernard had talked first? Or, what if, instead of taking turns, only one of them was allowed to talk? Or, what if there were restrictions on what they could communicate? In general, one can wonder the following: what would have been the outcome of the interaction if the conversation had a different structure? This shows the importance of the communication's *protocol*: the established sequence of actions through which the agents will interact. In the birthday example, the protocol that the conversation between Albert an Bernard follows (Albert starts and then they take turns) is successful. Are there other protocols that are successful in this scenario? Would this protocol be successful in a different situation? Questions like this are relevant in computer science, in particular in communication within computer networks.[10] If one is interested in this kind of questions, one might be also interested in a more fundamental issue. In the example, the initial information that Cheryl provided, together with the conversation Albert and Bernard had, allowed them to get to (commonly) know a specific truth: Cheryl's birthday is on July 16^{th}. But, what about other facts? Is their initial information and communication protocol enough for them to get to know *everything* that is relevant to the situation? More generally, and taking a philosophical perspective, can an agent get to know all true sentences? In philosophy, advocates of the *verificationist thesis* answer this question positive, and yet opposite arguments as Frederic Fitch's *paradox of knowability* have been proposed. In the epistemic logic setting, it turns out that not all truths can be known. An agent might not know that a certain fact "p" is true, thus making the sentence "p is true and she does not know it" also true. However, there is no way for her to (irrevocably) know that "p is true and she does not know it". If she were told that the sentence is true, then she would learn its first part: p holds. But then, she cannot know its second part, that she herself does not know p: now she knows it. This shows how these formal tools can shed new light on philosophical issues.

[10] For example, the *Russian card problem*, discussed in cryptography-related circles, poses the following question: can two agents share information in order to make a given secret common knowledge among them, while still making sure that a third agent, eavesdropping on the communication, does not get to know the secret?

5 • Further Perspectives

While the restaurant example shows how informational actions are not restricted to acts of inference, the analysis of the birthday example sketches how 'complex' real-life scenarios of multi-agent interaction can be modelled with logical tools. These examples are meaningful for modern logic, which can be understood as the study of *(i)* the different information-changing *actions* that real and/or artificial *agents* (e.g., human beings, but also computer programs) can perform, *(ii)* the way these actions *affect* the agents' individual and collective *information*, and *(iii)* the way these actions *interact* with one another, creating in this way long-term information-changing procedures.

These examples also show how this broader view on logic establishes meaningful connections with other fields, thus raising further interesting questions about the way information flows in multi-agent systems. First, the discussion on non- valid inference and the attempts to model forms of human reasoning connect logic not only with artificial intelligence, but also with psychology and cognitive science. The latter are particularly useful not only when looking for patterns of inference that are used by humans, but also to explain why other patterns are not. When factors that play a role are identified (e.g., source of information, emotions, moral values), one can incorporate them into a logical formalism, thus providing a more faithful descriptive model of the way humans reason.

Then, the further discussion on the birthday's example highlights some questions that are also important in philosophy and computer science. It has been already argued why individual agents cannot get to know everything that is true (see the end of previous section). But then, can we characterize the kind of truths that an individual agent can get to know? Equally important: how? That is, which are the sequences of actions that lead to this outcome? Of particular importance: can agents get to know what they *distributively* know? If so, can they further *commonly* know everything they distributively know? And, if so, again: how? These are not the only connections this broader view of logic allows for. As simple as it is, the restaurant example can already be understood as a scenario of *game theory*, a powerful mathematical tool used for analyzing competitive situations in which the outcome of the 'game' (e.g., how many 'points' each agent will get) depends not only on the action an individual agent chooses, but also on the choices of the other agents. Indeed, as discussed, the waiter had several options for her questions to the family (e.g., sincere or insincere), and so did the family for their answers (truthful or deceitful, public or private). They made their choices according to the common goal of having the dishes distributed correctly, but had their goals been changed, their actions might have varied, and the outcome could have been different. In scenarios like this, an information- theoretic and logical perspective has added value. By allowing a representation of the potentially different knowledge, beliefs, preferences and goals the agents might have, it allows us to focus not on what each agent should do given the situation, her preferences and her available actions, but rather on what each one of them can do given her *knowledge* about the situation, the *information* she has about everybody's preferences, and the actions she *considers* are available (van Benthem et al., 2011; van Benthem, 2011/2014).

Another natural connection that can be made is with the social sciences. We provided the reader with different examples which illustrate that understanding the epistemic social environment in which we live and operate is crucial for the success of our individual daily decisions and actions. The examples we gave are not straightforward, as the ongoing discussion on social platforms reveal: a quick and superfluous analysis of the Cheryl's birth puzzle may lead you straight to the wrong answer. Yet, one can learn how to analyze one's own information state as well as the information state of others. Similarly, one can learn how to analyze *the explicit and implicit* information present in the social environment we operate in. In our presentation of the restaurant example and Cheryl's birthday scenario, we did focus mainly on the *explicit* exchange of information: questions are asked and answers are provided, and statements are publicly communicated. But there are also scenarios in which our beliefs, preferences and opinions are affected by more subtle acts, and even only by the (actual, imagined or implied) presence of others. This phenomenon, called *social influence*, can take many forms, as socialization (inheriting and disseminating norms, customs, values and ideologies), conformity (changing attitudes, beliefs and behaviours to match those of the majority), compliance (changing favourably in response to explicit or implicit requests made by others), reactance (adopting a view contrary to what the person is being pressured to accept) and obedience (changing in response to a direct command from an authority figure). This perspective opens up a new line of work in which logic can be used to study informational processes in social systems, including the study of different social epistemic phenomena, such as informational cascades (a number of people make the same decision in a sequential fashion) and pluralistic ignorance ("no one believes, but everyone thinks that everyone believes").

We highlighted the richness of logic and the usefulness of its tools to model different types of scenarios. However, even while the examples in this chapter were all analysed by making use of small-size models of low-complexity, it is important to note that the full power of our tools will become more apparent when the complexity of the models increases by adding agents, propositions, actions and by expanding on the type of attitudes.[11] Overall, when equipped with the tool-kit described in this chapter, ranging from classical logic all the way to non-classical logics, the reader will be ready to start navigating the spaces in which different informational processes appear. You will find the logical tools to be a crucial ingredient when operating in a variety of different contexts, including in our digital-social environments which have become increasingly more complex and allow for new forms of information-exchange every day.

[11] We refer the interested reader to Baltag & Moss (2004) and Baltag et al. (2008) for further information about different such scenarios. For readers who are new to logic, a systematic study will require to get first acquainted with the basics of classical logic, followed by a study of modal logic in order to construct the type of possible worlds models we have informally sketched in this chapter. We refer to van Benthem (2010) for an introduction to the required tools of modal logic, after which the reader can proceed with a study of dynamic epistemic logic. The entry in Baltag & Renne (2016) provides an excellent overview of the state of the art on dynamic epistemic logic and supplements the book van Ditmarsch et al. (2015) on this topic.

BIBLIOGRAPHY

Baltag, A., & Moss, L. (2004). Logics for epistemic programs. *Synthese*, *139*(2), 165–224. 19

Baltag, A., & Renne, B. (2016). Dynamic epistemic logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. 13, 19

Baltag, A., van Ditmarsch, H., & Moss, L. (2008). Epistemic Logic and Information Update. In P. Adriaans & J. van Benthem (Eds.), *Philosophy of information* (pp. 361–455). Elsevier. 19

Fagin, R., Halpern, J., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. The MIT Press. 9

Shields, C. (2016). Aristotle. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Stanford University Press. 2

van Benthem, J. (2010, April). *Modal logic for open minds*. CSLI Publications. 19

van Benthem, J. (2011). Logical dynamics of information and interaction. *Logical dynamics of information and interaction*. Cambridge University Press. 10, 18

van Benthem, J., Pacuit, E., & Roy, O. (2014). Toward a theory of play: A logical perspective on games and interaction. *Games*, 2(1), 52–86. (Original work published 2011) 18

van Ditmarsch, H., Halpern, J. Y., van der Hoek, W., & Kooi, B. (Eds.). (2015). *Handbook of epistemic logic*. College Publ. 9, 19

CONTACT

Sonja Smets s.j.l.smets@uva.nl
Fernando R. Velázquez Quesada f.r.velazquezquesada@uva.nl

CITE AS

Smets, S., & F. Velázquez Quesada (2021). The study of informational processes: An introduction to logic. In M. Sie & B. Engelen (Eds.), *Second thoughts: First introductions to philosophy*. Open Press TiU.
https://doi.org/10.26116/secondthoughts-sie-engelen-openpresstiu-2021-smets-velazques-quesada-1.

LICENSE