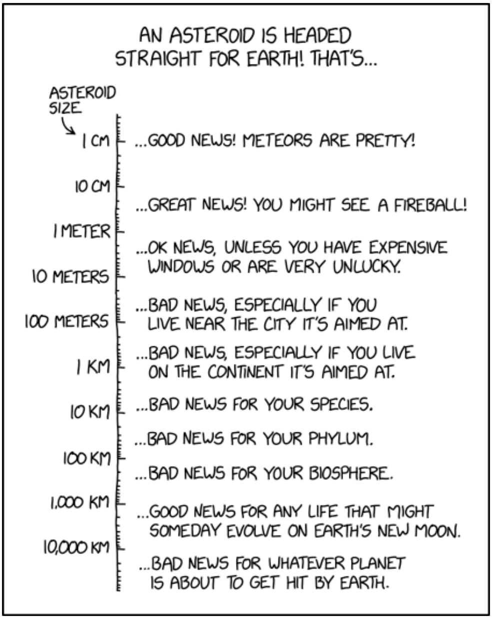
The Reasoner

Volume 19, Number 2
April 2025

ISSN 1757-0522
Milano University Press
riviste.unimi.it/index.php/thereasoner


AN ASTEROID IS HEADED STRAIGHT FOR EARTH! THAT'S...
ASTEROID SIZE
1 CM
10 CM
1 METER
10 METERS
100 METERS
1 KM
10 KM
100 KM
1,000 KM
10,000 KM
...GOOD NEWS! METEORS ARE PRETTY!
...GREAT NEWS! YOU MIGHT SEE A FIREBALL!
...OK NEWS, UNLESS YOU HAVE EXPENSIVE WINDOWS OR ARE VERY UNLUCKY.
...BAD NEWS, ESPECIALLY IF YOU LIVE NEAR THE CITY IT'S AIMED AT.
...BAD NEWS, ESPECIALLY IF YOU LIVE ON THE CONTINENT IT'S AIMED AT.
...BAD NEWS FOR YOUR SPECIES.
...BAD NEWS FOR YOUR PHYLUM.
...BAD NEWS FOR YOUR BIOSPHERE.
...GOOD NEWS FOR ANY LIFE THAT MIGHT SOMEDAY EVOLVE ON EARTH'S NEW MOON.
...BAD NEWS FOR WHATEVER PLANET IS ABOUT TO GET HIT BY EARTH.

The bottom ones are also potentially bad news for any other planets in our solar system that have been counting on Earth having a stable orbit

https://xkcd.com/3049/

The Reasoner 19(2), April 2025.

Contents

Guest Editorial **1**

Agent-based models in philosophy
:: Lorenzo Casini 1

Features **2**

An Interview with Dunja Šešelja, Samuli Reijula, and Matteo Michelini
:: Lorenzo Casini 3

Research Articles **25**

Disconfirmation is not modus tollens
:: Kenneth Aizawa 26

Post-publication reviews and comments **32**

Lack of experience is a real danger in the diagnostic process
:: Bimal Jain 33

The Dissemination Corner **35**

Interdisciplinary Systematic Review: A Novel Approach to Evidence Synthesis
:: Sahanika Ratnayake 36

Book Reviews **39**

M. D' AGOSTINO, S. MODGIL AND C. LARESE. *Depth-Bounded Reasoning. Volume I: Classical Propositional Logic. College Publications, London. 2024, xvii + 225, ISBN 978-1-84890-442-2*
:: ALEJANDRO SOLARES-ROJAS 40

Agent-based models in philosophy

:: Lorenzo Casini

Abstract
The Guest Editor introduces the relevance of Agent-based models in Philosophy, the topic of this issue's opening feature

Keywords
Agent-based models; philosophy of science

How to Cite
Casini, L. Agent-based models in philosophy: Guest Editorial. The Reasoner 19(2). https://doi.org/10.54103/1757-0522/28586

I am very glad to be again Guest Editor for *The Reasoner*. This time, my chosen topic is agent-based modelling (ABM) in philosophy of science. Agent-based models (ABMs) have revolutionized our approach to understanding complex systems across numerous fields. Unlike traditional analytical methods that often rely on simplifying assumptions and struggle with non-linear relations, ABMs simulate from the bottom up the behaviour of individual, heterogeneous agents and their interactions. ABMs have been used to address critical questions in epidemiology (e.g.

http://doi.org/10.54103/1757-0522/28586

spread of disease through social networks), economics (e.g. emergence of financial market bubbles), and the social sciences (e.g. dynamics of opinion polarization), among other disciplines.

In the last few years, following their success in modelling complex systems across scientific disciplines, ABMs have become the weapon of choice of more and more philosophers of science, too, for investigating phenomena such as theory choice, scientific collaboration, and knowledge diffusion. Without further ado, let us delve deeper into the topic with Dunja Šešelja, a leading expert in the field, and two of her collaborators, Samuli Reijula and Matteo Michelini.

Lorenzo Casini

University of Bologna

An Interview with Dunja Šešelja, Samuli Reijula, and Matteo Michelini
:: Lorenzo Casini

Abstract
This interview discusses the use of agent-based modelling in philosophy of science with leading scholars in the field, Dunja Šešelja, Samuli Reijula, and Matteo Michelini.

Keywords
Agent-based modelling; computational social science; philosophy of science.

How to Cite
Casini, L. An Interview with Dunja Šešelja, Samuli Reijula, and Matteo Michelini. The Reasoner, 19(2). https://doi.org/10.54103/1757-0522/28587

Dunja Šešelja is Professor for Social Epistemology and Reasoning in Science at the Institute for Philosophy II, Ruhr-Universität Bochum, and a member of the research group Reasoning, Rationality and Science*. She is also the Initiator of the DFG Scientific Networks Grant on* Simulations of Scientific Inquiry*, a project she's going to describe in more detail during our interview.*

Lorenzo Casini: Hi Dunja, thank you for accepting to be interviewed for *The Reasoner*. ABMs have become a powerful tool for

http://doi.org/10.54103/1757-0522/28587

understanding complex systems. Can you tell us how you came to use them in your research? What questions were you asking at the time, and how did this tool help you answer them? Could you provide a concrete example of a philosophical problem that you found particularly well-suited for an ABM approach as opposed to a more traditional philosophical method?

Dunja Šešelja: Thanks for having us, Lorenzo! To answer your question, let me start with a bit of background. My PhD at Ghent University was at the intersection of formal studies of scientific controversies—using formal theories of argumentation—and integrated history and philosophy of science. Around that time (early 2010s), ABMs were gaining traction in the social epistemology of science, particularly through the work of Kevin Zollman, Michael Weisberg and Ryan Muldoon, Igor Douven, and my colleague at Ghent University, Rogier De Langhe. These models introduced provocative and counterintuitive insights about the social dynamics of scientific inquiry. For instance, they suggested that too much information flow could sometimes hinder collective inquiry—what came to be known as the "Zollman effect" after Kevin's work—or that "sticking to one's guns" in scientific disagreements might, under certain conditions, be beneficial for the community of scientists.

What drew me to ABMs was their ability to illuminate complex, emergent social dynamics that traditional philosophical methods struggle to capture. While traditional methods allow us to ask, "What is a rational response to peer disagreement?" ABMs enable us to ask a different, yet equally important question, "If individual scientists respond to a disagreement in a certain way, how will that impact their collective inquiry?" This shift from individual rationality to complex social dynamics is crucial for understanding the social dimension of science.

At the same time, I was skeptical of taking the results of highly idealized models at face value without scrutinizing their robustness. From the philosophy of modeling literature, it was clear that different types of explanations emerge from abstract and highly idealized models depending on their level of validation. Robustness checks—such as testing the stability of results under parameter variations or changes in structural assumptions—here play an important role. For example, if we increased the number of agents, or if agents in the model conduct their research in a slightly different way, would we still get the same result? If a result is highly sensitive to these factors, then we at least gain insight into its scope and limitations. In this way, we can not only generate surprising hypotheses but also zoom in on the specific conditions under which those hypotheses hold.

My initial engagement with ABMs was therefore motivated by a desire to critically assess the robustness of existing models and explore how their insights could be made relevant to the broader philosophical community, beyond the niche of formal social epistemology. As I was already interested in the argumentative dynamics underlying scientific inquiry during my PhD, I decided to take a further step and collaborate with others to develop an argumentative ABM of scientific inquiry, which served to test the robustness of previously proposed models.

LC: Transitioning from traditional philosophical methods to ABM often presents unique challenges. Could you describe the key hurdles you and other philosophers faced in adopting this approach, particularly in acquiring the necessary technical skills? And did you find it challenging to translate abstract philosophical concepts, like “theory choice” or “scientific collaboration”, into concrete, operationalized models?

DS: Since I don’t have a background in programming, one of the key challenges in adopting ABMs was acquiring the necessary technical skills. Fortunately, I was surrounded by collaborators who did have this expertise, and I was able to learn a great deal from them. My three main collaborators back then—AnneMarie Borg, Daniel Frey, and Christian Straßer—all had programming experience, which made it much easier to get started. I also benefited from exercises on NetLogo, a programming language commonly used for ABMs, which were originally developed by Conor Mayo-Wilson while working at the Munich Center for Mathematical Philosophy (MCMP), LMU Munich, and shared with me by my colleagues during my postdoctoral time there.

That said, learning to code for ABMs is much more accessible today than it was when we first started. Newer programming languages, such as Julia, offer alternatives that are both powerful and relatively easy to pick up. Moreover, ABM is increasingly taught to philosophers, too. For instance, in our research group in Bochum, Matteo Michelini teaches a course on NetLogo, which has been a great resource for students and researchers interested in this approach.

Beyond the technical hurdles, another major challenge with ABM is not just translating philosophical concepts (such as theory choice or scientific collaboration) into formal models, but also identifying research questions that are well-suited to an ABM approach. When conceptualizing a model, it’s important to avoid two extremes—modeling a question that could be answered without simulation, and tackling a question so complex that an abstract model cannot provide a clear or interesting answer. Typically, this means focusing on how specific activities of individual scientists impact collective inquiry, while abstracting away from numerous other factors that influence real-world science. This approach allows us to ask “what-if” questions that shed light on the social dynamics of scientific inquiry.

For example, traditional philosophical studies of theory choice have focused on the criteria under which individual scientists rationally decide which theories to pursue, or which theories to accept. However, this leaves open the question how different standards of theory choice impact collective inquiry. For instance, is a scientific community better off if its members favor pursuing theories that have a larger explanatory scope than the rivals, or should scientists instead remain committed to their current theories—even if their scope is more limited—until they suffer from too many anomalies in comparison to the rivals? An ABM allows us to explore how a scientific community that prefers one strategy over the other performs over time, offering insights into the consequences of different inquisitive norms on the epistemic performance of the group.

LC: Introducing a novel methodology like ABM can elicit varied reactions within a well-established discipline. How did the philosophical community initially react to this new approach? Was it difficult to navigate philosophers’ expectations and demonstrate the rigor and significance of your findings in a way that resonated with more traditional philosophical standards?

DS: As I mentioned earlier with reference to my personal skepticism in approaching the method during my PhD, if we want the broader philosophical community to take the results of our models seriously, it is crucial to clarify their epistemic function. A common objection to highly idealized ABMs is that they omit many factors that play an important role in actual scientific inquiry. However, this critique often stems from a misunderstanding of what these models are meant to show. In most cases, ABMs are not designed to explain or predict real-world scientific inquiry directly. Instead, they may show how certain patterns or phenomena can emerge from a minimal set of conditions—some of which we might not have previously considered.

For instance, a model might show that a scientific community can become polarized even if all its members are individually rational, provided they start with different background assumptions. The question is not whether real-world polarization is necessarily caused by this specific mechanism but rather whether such a mechanism is capable of producing polarization. Of course, in some cases, ABMs can also be empirically embedded to investigate whether particular socio-epistemic factors played a role in historical episodes of scientific inquiry. But this requires additional work to connect the model to empirical data.

LC: The field of ABM in philosophy of science has seen significant growth in recent years. Can you trace a map of the different directions in which the ABM community has grown? What results have been achieved and what lies ahead in terms of challenges and unanswered questions?

DS: Yes, the ABM community in the philosophy of science has grown significantly, both in terms of modeling frameworks and research topics. A substantial body of work now exists on network epistemology, the division of cognitive labor, the conditions that are conductive to theoretical diversity in science, the role of cognitive and social diversity in collective inquiry, factors driving scientific polarization, and so forth.

Looking ahead, several avenues remain underexplored. First, most ABMs developed by philosophers are still largely abstract and not empirically informed. While abstraction is not necessarily problematic, an open question remains: Can some of these models be empirically embedded to provide normative insights, for instance, guiding interventions in specific contexts of inquiry? Second, ABMs have been extensively developed in other disciplines, such as the social sciences and computer science, and modeling frameworks emerging in these other disciplines could be fruitful in addressing philosophical questions. Third, the recent surge in machine-learning research presents an exciting, yet underexplored, opportunity to integrate ABMs with machine-learning approaches, potentially offering new ways to study the dynamics of scientific inquiry. For instance, Gregor Betz has started using such approaches to introduce natural language into argumentative ABMs (see here).

LC: Collaborative networks play a crucial role in advancing academic research. You're the Initiator of the DFG Scientific Networks Grant on *Simulations of Scientific Inquiry*. Can you tell our readers what the purpose of the network is and what are its plans for the near future?

DS: Our network was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in 2019, during my time at the MCMP. Originally, we planned to hold our first workshop in April 2020, but with the outbreak of the COVID-19 pandemic, all in-person events had to be canceled. In response, we launched a series of online sessions featuring both network members and guest speakers presenting their work. This online series turned out to be a great success, with presentations not only from network members such as Patrick Grim, Christoph Merdes, Cailin O'Connor, Samuli Reijula, and Kevin Zollman, but also from guest speakers such as Igor Douven, Sina Fazelpour, Reiner Hegselmann, Paul Smaldino and Leonid Tiokhin.

After three years of virtual meetings, we were finally able to organize our first in-person event in February 2023. By that time, the number of junior scholars working on ABMs had grown significantly, and our workshops brought together a fantastic mix of PhD students and senior scholars from around the world. Over the next two years, we organized four major events featuring keynote speakers who are experts in ABMs (Gregor Betz, Ulrike Hahn, Toby Handfield, Rainer Hegselmann, Conor Mayo-Wilson, Cailin O'Connor, Erik Olsson, Samuli Reijula, Patricia Rich, Hannah Rubin, and Kevin Zollman), as well as philosophers working on the epistemology of modeling (Wybo Houkes and Paul Hoyningen-Huene) and scholars in formal philosophy of science using other formal approaches (including our interviewer, Lorenzo Casini, and Finnur Dellsén).

One of the most remarkable aspects of these events was the continued presence of PhD students, who became regular participants even though they were not officially part of the network. Additionally, inspired by our initial online series during the pandemic, we launched the *Computational Social Philosophy Seminars*, which continue to foster international interaction and collaboration.

Given the growth of the ABM community in recent years, we are now preparing a new research network grant application to formally include the next generation of scholars. This new network will expand beyond ABMs of science to computational social epistemology, focusing on emerging trends such as argumentative ABMs, computational studies of epistemic democracy, and computational analyses of epistemic injustice.

*

My second interviewee is Samuli Reijula. Samuli is affiliated to the University of Helsinki and a member of the Centre for Philosophy of Social Science (TINT). He is currently an Academy-of-Finland Research Fellow, leading a project titled "Modeling the Republic of Science: Collaborative Problem Solving and Collective Rationality in Scientific Inquiry". He is also a member of the DFG network on Simulations of Scientific Inquiry *led by Dunja.*

LC: Thank you for joining us, Samuli! Your research focusses on the nature of scientific problem solving at different levels, from individual scientists to entire scientific communities. Could you please elaborate on what you mean by "well-functioning scientific research"?

Samuli Reijula: Thanks! I think the goal of the social epistemology of science has been nicely expressed by Philip Kitcher and by Steve Fuller. According to Kitcher (*The Advancement of Science*, Oxford University Press, 1993), "the general problem of social epistemology is to identify the properties of well-designed social systems" (p. 303). Fuller gives a bit more detail:

> How should the pursuit of knowledge be organized, given

> that under normal circumstances knowledge is pursued by many human beings, each working on a more or less well-defined body of knowledge and each equipped with roughly the same imperfect cognitive capacities, albeit with varying degrees of access to one another's activities? (*Social Epistemology*, Indiana University Press, 1988, p. 3)

Fuller's characterization highlights some of the key factors that feature in the models we often use: diversity in cognitive resources of boundedly rational agents, and the communication patterns between them. I think as philosophers of science, we're ultimately interested in the normative question of how the practices of science should be organized so that the production of knowledge is efficient, reliable and fair. How each of those values should be understood, is, of course, a topic on its own.

In my own modeling work, I often adopt the perspective of viewing science as a distributed problem-solving machine, "problem-solving writ large". From the distributed-cognition perspective, each researcher can be seen as a component of a larger (eco)system. I don't want to overlook the significance of the epistemic feats of individual scientists, but I think it's fair to say that most of the genuinely interesting and important problems we face as humans exceed the cognitive resources of any isolated individual. It is only together, as research groups, collectives, research fields, and finally, as "the" scientific community, that we can hope to address those challenges.

I think that for understanding such distributed systems, ABMs can be a very useful tool. Like Dunja suggested, even when we work with simple assumptions about how individual agents behave and combine those with further, simple assumptions about the institutions of resource allocation, communication, division of cognitive labor in the scientific community, we can sometimes come up with interesting insights about how the problem-solving machine of science works, and how it could work.

LC: Scientific collaboration is increasingly recognized as a key factor in successful research. What led you to focus on collaborative problem solving? Could you explain the kind of computational models you are using, and what they help illuminate?

SR: Ever since I encountered Herbert Simon's theory of *problem solving as search* in an undergraduate cognitive science course, I've been very inspired by the idea of scientific discovery as heuristic search.

Problem solving—and the creative discovery it often requires—is in some important ways different from "mere" decision making. Unlike in traditional models of decision making, in problem solving the set of potential solutions is typically not known beforehand, the candidate solutions must be discovered. For example, think about the problem of finding a chemical substance that would work as an antibiotic against a superbug. Since the abstract space of possible chemical compounds is large, how is one supposed to search in there?

By the way, the idea of representing a problem structure metaphorically as a high-dimensional space has recently become concrete reality in automated labs doing drug discovery and materials science. Such "self-driving labs", as they are called, explore spaces of possible chemical compounds or new materials by combining optimizing algorithms with automatically run series of material experiments. Our traditional models of scientific experimentation suggest that we test hypotheses one at a time, but these systems are massively parallelized and can quickly sift through thousands of hypotheses. It's wild.

There are a couple of reasons why I think heuristic-search models are particularly fitting for studying scientific problem solving. As I have argued together with my friend and colleague Jaakko Kuorikoski (see here), in order to study collaborative problem solving in science, the modeling approach must be capable of representing some key features of the process, namely (1) problem structure and problem decomposition, (2) diversity of cognitive resources, and (3) cognitive coordination between problem-solving agents.

I think that the models in biology and organization science, inspired by Simon's early work (see, e.g., work by Stuart Kauffman, Jim March, and Daniel Levinthal), provide a good starting point for developing models of scientific problem solving, as shown by recent work in philosophy of science (see here, here, and here).

LC: You've explored the trade-off between "transient" diversity and diffusion of ideas in group problem solving. Could you explain what this trade-off is, and what your simulations suggest about how to balance these factors for better problem solving?

SR: Cognitive diversity has been shown to trade-off against a couple of things. The first is the diffusion of ideas in a population of agents as examined in detail in network epistemology models, first introduced in Kevin Zollman's seminal "The Communication Structure of Epistemic Communities, Philosophy of Science" (*Philosophy of Science*, 74(5), 2007) and "The Epistemic Benefit of Transient Diversity" (*Erkenntnis*, 72, 2010) papers. There, Kevin derives the *Zollman effect* that Dunja mentioned before. It's this counter-intuitive finding that theory choice in a community of scientists may be more reliable when the agents communicate less with each other.

The second sort of trade-off, the one that I've examined the most in my research, concerns the tension between group diversity and individual expertise. According to a famous finding by Lu Hong and Scott Page ("Groups of diverse problem solvers can outperform groups of high-ability problem solvers", *PNAS*, 101(46), 2004), called *diversity-trumps-ability theorem*, a group of diverse problem solvers can outperform a group consisting of individually best performing problem solvers. Although I've shown (see again here) that there are technical problems with the original finding, I think the basic idea is solid and important to the social epistemology of science: Looking at scientific progress we're often too focused on the role of exceptional individuals or geniuses. A more useful perspective on scientific progress is to look at the cognitive resources—research questions, solutions to problems, methodologies—distributed between the different members of the scientific communities.

It seems that discoveries and breakthroughs happen when those ingredients—you can imagine them floating around in the intellectual space of the scientific community—are brought together in the problem-solving activities of a diverse research group. Diversity can beat ability because high-ability experts often tend to resemble each other, and consequently, a group of high-performing experts ends up having access to fewer cognitive resources.

LC: Given your focus on institutional design, how do you envision your research informing practical strategies for improving scientific collaboration and problem solving in real-world settings? What are its possible implications for funding agencies, research institutions, and individual scientists?

SR: Almost every extended discussion on ABMs eventually turns to this question! And it's a really important question. I tend to fall on the side of caution here: We should not oversell the policy-relevance of the findings coming from theoretical models. My caution comes from the work I've done on evidence-based policy. From the policy world, we know how difficult it is to apply theoretical insights to real-life situations—this is the well-known problem of extrapolation.

That said, I think the policy world might have swung too far in the direction of inductive skepticism. The core idea of evidence-based policy has been to replace theorizing with experimentation at target site, so as to avoid the risky inductive generalizations between different populations. But I don't think we can do without theory.

Research policy is an example of an area where models and data could work together: Modeling can help us uncover mechanisms, or at least abstract motifs, that can be detected across several empirical situations (e.g. cumulative advantage, belief polarization, transient diversity). Especially in complex situations where many of such mechanisms interact and trade-off against one another, our intuitions become unreliable, and therefore, we need models to understand the dynamics.

So, to answer your question, I don't think these models alone have policy implications, but I do think that modeling should often be a part of the process of research-based policymaking. ABMs, in particular, are a powerful way of bringing together theoretical ideas, assumptions, and data, and exploring what follows from their conjunction.

I think that in the fields of *meta-science* and of *science of science* there are interesting examples of how to combine theory with empirical data for thinking about research policy, but I also think both of these research fields would benefit from more engagement with the philosophy of science.

LC: Looking ahead, what are the most pressing unanswered questions in the study of scientific collaboration and problem solving, and how do you see your research contributing to future advancements in this field?

SR: For a long time I was an AI skeptic, and even after the November 2022 introduction of LLM chatbots, I had decided not to jump on the AI-philosophy bandwagon. But somehow I started reading up on DeepMind's AlphaFold, namely the AI system that practically solved the protein folding problem, and self-driving labs.

Advocates of these approaches claim that automated discovery methods will revolutionize the scientific method in the coming years. That's of course a marketing pitch, but I think it is beyond doubt that now we do have existence proofs of automated scientific discovery. AlphaFold's protein structure library has influenced the field of biomedical research in a significant way. I've now been applying for funding for a project to study scientific problem solving and discovery in AI-enhanced labs.

Looping back to Herbert Simon whom I mentioned above, I think that concrete examples of automated discovery systems like AlphaFold could be what he always dreamed of: Simon thought scientific discovery should be understood as continuous with everyday problem solving, and that it could be explained in terms of the models we have in cognitive science. Automated discovery systems could be a new source of evidence for that kind of theory: despite being huge in terms of data and computational capacity, their fundamental architecture is understandable and fully transparent as program code.

So maybe the emergence of AI discovery systems could stage a comeback for the 1980s agenda of the "friends of discovery", who aimed to make of the study of scientific discovery an integral part of the epistemology of science.

*

My third interviewee is Matteo Michelini, a cotutelle PhD student at Ruhr-Universität Bochumand TU Eindhoven, under the supervision of Dunja Šešelja and Wybo Houkes.

LC: Hi Matteo. After completing a Masters in Logic, during your PhD you turned to applying ABM to philosophical problems. What drew you to this approach? What do you think it uniquely contributes to social epistemology and philosophy of science?

Matteo Michelini: What drew me to ABMs? The fact that I *hated* logic. Of course, I'm joking! Still, on reflection my shift towards ABM was indeed grounded in three differences between simulation tools and analytic/logical tools.

First, I've always been attracted more to theories that explain and assess social phenomena, as opposed to the study of languages, or knowledge per se. It is something I developed when I was doing my Bachelor in Philosophy studying philosophers like Locke, Rosseau, or Marx.

Second, during my Masters, I did not just study logic but also game theory and computational social choice, which are used in computational simulations. These frameworks fascinated me. Given my background— I started out as a Mechanical Engineering student, I confess—I saw them as an exciting way to apply my math skills—and make my first years of study fit better in my life narrative—good choice after all, wasnt' it?

Finally, I also liked the idea of *improving* society—whether scientific communities or broader groups navigating epistemic challenges. My family upbringing instilled in me a strong sense of social responsibility, so this naturally resonated with me.

At some point, I just thought, *wow, I can use the formal tools I love to answer questions I care about—and even tell myself they have some societal relevance.* That said, I quickly realized that achieving this last goal wouldn't be so straightforward—but I suppose that's a topic for another question.

As for the unique contribution of ABMs, the answer is complex, as Dunja already made quite clear. In fact, their epistemic role is subject of an ongoing discussion within the community. Personally, I like to think of them as thought experiments made precise, computational tools that allow us to rigorously derive the consequences of our assumptions (especially if these assumptions are simple—something which both Dunja and Samuli referred to). As such, they are the perfect tools to study how social dynamics may bring about certain phenomena, or test explanations for collective behaviour. In short, ABMs enable us to reason about collective phenomena in a precise and grounded way—-without blurred lines or half-sketched consequences.

LC: Your past research explores how epistemic practices detrimental at the individual level can negatively affect the scientific community. Currently, you're investigating how these very same practices may also—surprisingly—turn out beneficial for the community as a whole. Could you share paradigmatic examples of this sort of phenomena?

MM: Sure, there are plenty of epistemic practices that are detrimental to individual inquiry but can have mixed effects at the collective level—beneficial under certain conditions and detrimental under others. In my work, I've primarily focused on two such cases, namely the effects of evidence misinterpretation in scientific inquiry and the role of individually vicious cognitive dispositions. For both of these, I first published papers outlining their potential downsides (see here and here), and I am now working on follow-up research exploring their positive contributions. As this is the most surprising aspect of my research, let me briefly explain how each of these mechanisms might actually enhance inquiry.

Evidence misinterpretation happens when someone draws the wrong conclusion from scientific evidence. Naturally, misinterpreting evidence is usually seen as something scientists should avoid at all costs. However, using ABMs, we've shown that—under very specific conditions—a moderate likelihood of misinterpretation can actually benefit the scientific community. The reason is that when scientists have a small probability of misinterpreting evidence, they are more likely to form diverse beliefs about which method is superior. This diversity leads to a more comprehensive exploration of available methods, ultimately increasing the chances of identifying the best one. This phenomenon is known as transient diversity, as Samuli has already explained.

A second example involves “myside bias” and laziness, respectively the tendencies to only produce arguments in favour of one’s opinion and to never critically evaluate such arguments before sharing them. My current research suggests that a group of biased and lazy reasoners might actually be more likely to reach the correct solution to a problem than a group of diligent and impartial ones, whenever cognitive and material resources are limited. The reason is that biased and lazy reasoners can sometimes better and faster leverage the collective competence of the group. But I’ll stop here—more details will be in my next paper, and I don’t want to spoil too much!

What’s important, though, is all these results hold under very specific conditions. Misinterpretation is beneficial only if each scientist is only very moderately inclined to misinterpret evidence; as soon as someone is highly likely to misinterpret evidence the scientific community will never reach consensus on the superior method. Myside bias and laziness are helpful only if reasoners start out with more or less balanced perspectives, and material and cognitive resources are limited. ABMs are ideal to highlight these conditions: they allow us to derive them from the outcomes coming from the interactions between individual agents.

LC: What are some challenges you’ve faced as a PhD student working at the intersection of philosophy and computational modelling? Any advice for others considering a similar path?

MM: Well, I’d say there are two main challenges in this career—one more fundamental, the other more contingent.

The first is a persistent feeling of not fully belonging anywhere. While most philosophers openly appreciate our work, I always get the impression that some don't see us as really doing philosophy. The same happens when interacting with social scientists: we seem to be doing too much theoretical modeling to be fully accepted as social scientists. Compare this to someone working in other, more traditional areas of philosophy—other philosophers might disagree with their views or even their methods, but no one questions whether they are doing philosophy. In my case, I often feel like I have to prove to philosophers that I belong. In a way, as a PhD student, I live with double imposter syndrome! s a result, I often find myself striving to prove the relevance of my results to social scientists and its philosophical bearing to philosophers.

The second challenge is that, since this is a relatively new approach, there's no clear roadmap for what you need to learn to become a good scholar in the field. Until now, I feel I've learned many different tools and yet I've only successfully applied a few. Did I learn them for nothing? To-be-determined. At the same time, I constantly feel like there are tools I haven't mastered even though I should— not because I didn't apply myself enough but simply because nobody ever said (or thought) I was supposed to learn them in the first place!

I imagine that someone working in, say, epistemic logic has a much more structured path. They learn the foundational theorems, the standard approaches, and the key techniques for carrying out proofs. In contrast, for those of us at the intersection of philosophy and computational modeling, the path is far less defined.

So, my advice? Try to connect with as many people as possible. Building connections helps you gain a sense of identity and a better sense of what's worth learning. In fact, I'd say the DFG network I am part of, *Simulations of Scientific Enquiry*, was incredibly important in both respects—and I'll never thank Dunja enough for bringing me in.

LC: Right. Being part of the DFG network must have given you exposure to a wide range of perspectives. How has this collaboration influenced your work? Are there any insights or directions you wouldn't have explored otherwise?

MM: There's no doubt that the DFG network has had a major influence on my research. Conversations with network members shaped my understanding of what methods are useful and how I should approach my work.

For example, during my Masters in Logic, I didn't work extensively with Bayesian methods and had no plans to use them in my PhD. Then, at my first DFG meeting, I realized that almost everyone was working with Bayesian epistemology. So I thought, *hmm, maybe there's something to this.* From that point on, Bayesian tools have become a central part of my work in some way or other. The same goes for many of the methods, approaches, and ideas I've used—most were heavily shaped by my interactions with the network members.

Another thing that made the network so valuable was how close-knit it was, bringing together both senior and junior members. This gave me the chance to observe how senior people handle things and to form my work habits in the image of great role models. Honestly, if it weren't for the DFG network, I don't think I would have been able to produce—I am not ashamed to say—such interesting results, or to learn so much about how academia works.

LC: Looking forward, if you could see one of your research findings tested in a real-world setting, which one would it be and why?

MM: That's a billion-dollar question. In principle, testing any of the findings I mentioned earlier could bring valuable insights. While our computational models produce fascinating and promising results, they primarily serve to generate hypotheses and propose possible explanations. They don't allow us to make concrete predictions or offer definitive explanations of real-world phenomena. So yes, empirically testing our results—and calibrating our models accordingly—is absolutely crucial.

If I had to pick one hypothesis to be tested in a real-world setting, I'd go with my work on bias and laziness—the idea that these cognitive tendencies might actually benefit reasoners engaged in deliberation, particularly in situations where evidence and time are limited. I see this as a strong candidate for future empirical testing for a few reasons.

First, the hypothesis is closely tied to cognitive science, a field with well-established empirical methods for testing claims about reasoning and group epistemic performance. In fact, similar claims about the benefits of certain practices in collective reasoning have already been explored within cognitive science.

Second, if validated, this research could have a meaningful societal impact. Right now, we teach students to suppress their biases—to actively counter their natural tendency to favor their own beliefs. We even design strategies to help with this: for example, encouraging students to "consider the opposite" to push back against their own bias and mental inertia. But what if, in certain contexts, bias and laziness actually improve group reasoning? If future studies support this, we might reconsider how we teach critical thinking. We could instead tell students that, in safe deliberative environments—where bias isn't likely to cause harm—there's no need to fight these tendencies. Instead, they can lean into them.

On the other hand, testing our models on scientific communities is much more challenging and would require significantly more groundwork. One major hurdle is the lack of a clear, universally accepted measure of epistemic success for real scientific communities. Without such a metric, it's difficult to assess the impact of different mechanisms on scientific inquiry. That said, I don't think this is an insurmountable problem. Developing a measurable notion of scientific success—one based on publication data and expert assessments—is something we should strive for. As it happens, I'm currently working on this with Eugenio Petrovich, a philosopher of science and scientometrician. Together, we hope to advance the current understanding of the (elusive) notion of scientific success, and consequently to deepen our understanding of "good" mechanisms of scientific inquiry.

Lorenzo Casini

University of Bologna

Disconfirmation is not modus tollens
:: Kenneth Aizawa

Abstract
Scientific disconfirmation has often been thought to be reasoning by *modus tollens*. This interpretation, however, misconstrues the conditionals in this scientific reasoning in terms of the material conditional, rather than in terms of causal conditionals. Scientific confirmation has also been thought to be a logical fallacy, affirming the consequent. Once one embraces the idea that scientists are reasoning in terms of causal conditionals, rather than the material conditional, we can avoid the peculiarity of the view that scientific confirmation is based on a simple logical fallacy. Interpreting scientists as reasoning about physical consequences of hypotheses enables a more charitable interpretation of scientific disconfirmation and confirmation.

Keywords
Disconfirmation; abduction; modus tollens.

Sumbitted 11 February 2025
Revised 20 March 2025
Accepted 28 March 2025

How to Cite
K. Aizawa, Disconfirmation is not modus tollens. The Reasoner 19(2). https://doi.org/10.54103/1757-0522/28223

Many philosophers of science have interpreted instances of scientific disconfirmation as instances of *modus tollens*. Karl Popper embraced this idea in his falsificationism: “The falsifying mode of inference here referred to—the way in which the falsification of a conclusion entails the falsification of the system from which it is derived—is the modus tollens of classical logic” (Popper, 2005, p. 89). Carl Hempel, discussing Ignaz Semmelweis’s reasoning about the causes of childbed fever, was similarly explicit about this analysis.

> If H is true, then so is I. . . .
>
> But (as the evidence shows) I is not true.
>
> H is not true.
>
> Any argument of this form, called modus tollens in logic, is deductively valid; that is, if its premisses . . . are true, then its conclusion . . . is unfailingly true as well. Hence, if the premisses . . . are properly established, the hypothesis H that is being tested must indeed be rejected. (Hempel, 1966, p. 31)

Imre Lakatos also embedded *modus tollens* in his account of the methodology of scientific research programs (Lakatos, 1978). Much more recently, Alexander Bird seems to have embraced disconfirmation as *modus tollens*: “Imagine a case where we are investigating a hypothesis h. From h plus an auxiliary hypothesis we deduce a testable proposition o. We devise a suitable experiment and find that o is false. Normally we would regard h as refuted” (Bird, 2022, p. 187).

Philosophers who interpret specific instances of scientific disconfirmation as *modus tollens* assume that the scientific conditional reasoning in disconfirmation involves the material conditional. To illustrate, consider a passage from Ignaz Semmelweis's discussion of the higher mortality rate due to childbed fever in his first clinic than in his second clinic:

> It has not been questioned and has been expressed thousands of times that the horrible ravages of childbed fever are caused by epidemic influences. By epidemic influences one understands atmospheric-cosmic-terrestrial changes ... by which childbed fever is generated in persons predisposed by the puerperal state. But if atmospheric-cosmic-terrestrial conditions of Vienna cause puerperal fever in predisposed persons, how is it that for many years these conditions have affected persons in the first clinic while sparing similarly predisposed persons in the second? (Semmelweis, 1983, p. 65).

Popper or Hempel would interpret Semmelweis's reasoning along the following lines:

> If the higher mortality rate in the first clinic is due to atmospheric conditions, then the first and second clinics will have the same mortality rate.
>
> It is not the case that the first and second clinics have the same mortality rate.
>
> Therefore, it is not the case that the higher mortality rate in the first clinic is due to atmospheric conditions.

The philosophical interpretation of Semmelweis's reasoning as *modus tollens* overlooks the familiar point that many conditionals in natural language are not material conditionals. Thus, one might say “If Jones eats all this ice cream, she will get sick.” This conditional presupposes some sort of causal connection between eating the ice cream and getting sick. Eating the ice cream will make Jones sick. By contrast, the material conditional does not have this implication. The material conditional is a mere truth function. “If I eat this ice cream, then the sun is 93 million miles from the earth,” while a false statement of causal connection, can be a perfectly good truth-functional material conditional.

Returning to Semmelweis’s comments, there is reason to think that he is thinking in terms of causal relations among things in the world. He mentions the epidemic influences *causing* the horrible ravages of childbed fever and atmospheric-cosmic-terrestrial changes *generating* childbed fever. In other words, Semmelweis is reasoning about what one might loosely call the physical consequences of atmospheric-cosmic-terrestrial changes, namely, that those changes would affect both hospital wards equally. But, upon observation, one finds that those physical consequences do not obtain. The material conditional of symbolic logic is not meant to capture the putative causal connections between the atmospheric-cosmic-terrestrial changes and childbed fever.

A further hint regarding Semmelweis’s thinking comes a few sentences later, when he claims that “epidemic influences cannot explain the differences in mortality” (Semmelweis, 1983, p. 66). This comment suggests the familiar thought that explanatory considerations have confirmation theoretic import. The thought is that the physical consequences of things in the world being as hypothesis H depicts them lead to things in the world being as E depicts them, but independent determinations reveal that things in the world are not as E depicts them. In other words, when scientists “deduce a testable proposition o”—as Bird would put it—what they are doing is determining the physical consequences of some hypothesis (typically in conjunction with various auxiliary hypotheses). This is a basis for disconfirmation. This, in outline, is what scientists, such as Semmelweis, mean when they imply that what a hypothesis "cannot explain" is disconfirming for that hypothesis. In other words, Semmelweis is thinking of disconfirmation as some sort of abductive reasoning regarding what a hypothesis cannot explain.

Should philosophers of science take scientists to be engaged in reasoning about causes in the world, about the physical consequences of things, they would be in a position to more charitably interpret scientific reasoning. Instances of hypothetical reasoning might be more charitably interpreted as involving something like causal conditionals, rather than the material conditional. Indeed, the idea broached for scientific disconfirmation might be carried over to scientific confirmation.

Recall that Semmelweis intended to support or confirm the view that "cadaverous particles" were responsible for childbed fever by supposing that, if the particles were chemically destroyed the mortality rate in the first clinic would be reduced: "Suppose cadaverous particles adhering to the hands cause the same disease among maternity patients that cadaverous particles adhering to the knife caused in Kolletschka. Then if the particles are destroyed chemically, so that in the examination patients are touched by fingers but not cadaverous particles the disease must be reduced" (Semmelweis, 1983, p. 65). Hempel might render this reasoning as an instance of affirming the consequent:

> If childbed fever is caused by cadaverous particles, then chemical cleaning reduces the incidence of childbed fever.
>
> Chemical cleaning reduces the incidence of childbed

> fever.
>
> Therefore childbed fever is caused by cadaverous particles.

What is uncharitable in this interpretation is that, as Hempel noted, it takes scientific confirmation to be a deductive fallacy. See (Hempel, 1966, pp. 31-32).

An alternative reading of Semmelweis's comments is that he is tracing the physical consequences of the hypothesis that childbed fever is caused by cadaverous particles. One physical consequence is that any chemical that is strong enough to kill the cadaverous particles would reduce the incidence of childbed fever. Subsequent investigation bears this out, thus supporting the idea that disease is caused by cadaverous particles. Semmelweis's underlying thought is that the chemical cleaning destroying the cadaverous particles explains why the cleaning reduces the mortality rate. In other words, Semmelweis is reasoning abductively.

The foregoing considerations support a familiar view, namely, that philosophers of science should abandon a hypothetico-deductive interpretation of disconfirmation and confirmation. Further, it suggests an alternative: at least some of the scientific reasoning that would formerly have been interpreted in a hypothetico-deductive framework might be better interpreted within an abductive framework that recognizes scientific thinking about the physical consequences of what hypotheses propose. An abductive interpretation links confirmation to explaining and disconfirmation to failing to explain.

*

Bird, A. (2022). *Knowing Science*. Oxford University Press.

Hempel, C. G. (1966). *Philosophy of Natural Science*. Prentice-Hall.

Lakatos, I. (1978). Falsification and the methodology of scientific research programmes. In J. Worrall & G. Currie (Eds.), *The Methodology of Scientific Research Programmes* (Vol. 1). Cambridge University Press.

Popper, K. (2005). *The Logic of Scientific Discovery*. Routledge.

Semmelweis, I. (1983). *The Etiology, Concept, and Prophylaxis of Childbed Fever* (K. C. Carter, Trans.). University of Wisconsin Press.

Kenneth Aizawa
Rutgers University

Lack of experience is a real danger in the diagnostic process

:: Bimal Jain

Abstract

A failure to suspect a disease with an atypical presentation and formulate it as a hypothesis is not due to heuristic of representativeness as proposed by Balzaretti but is due to lack of awareness of atypical presentations of a disease due to lack of experience. Thus, lack of experience is a real danger in the diagnostic process.

Keywords

Heuristics; diagnosis; experience.

How to Cite

Jain, B. Lack of experience is a real danger in the diagnostic process. The Reasoner, 19(2). https://doi.org/10.54103/1757-0522/28117

Dear Editor,

I do not agree with several comments made by Balzaretti (Representativeness heuristic is a potential danger to the diagnostic process. The Reasoner. 19(1). https://doi.org/10.54103/1757-0522/27372) about my article (Jain BP. 2024. Role of heuristics in diagnostic reasoning in practice. The Reasoner. 18(4): 32). First of all, his comment that diagnostic reason-

http://doi.org/10.54103/1757-0522/28117

ing is similar to ordinary, everyday reasoning is wrong. Diagnostic reasoning, like all scientific reasoning, is well-known to be a process of hypothesis generation and testing (Jain BP. The scientific nature of diagnosis. Diagnosis 4 (1):17-19. doi: 10.1515/dx-2016-0032), which is lacking in everyday reasoning. The role of heuristic of representativeness in diagnostic reasoning, as I mention in my article, is to make us suspect a disease from a presentation and formulate it as a hypothesis, which is then tested and diagnosed (or not) after testing. In everyday reasoning, on the other hand, this heuristic leads to a probability judgment directly from available information without any hypothesis generation and testing, that is usually erroneous, as we see in the engineer-lawyer experiment (Jain BP. 2024: 32). Secondly, Balzaretti argues this heuristic prevents us from suspecting a disease with an atypical presentation by citing example of a 70 year old woman in whom aortic dissection was not suspected, due to its atypical presentation. I believe it was not suspected due to this heuristic but due to lack of awareness of wide range of presentations of aortic dissection including those that are atypical, due to lack of experience. We find experienced physicians to routinely suspect diseases with atypical presentations and diagnose them accurately after testing , in scores of published diagnostic exercises about real patients such as in clinical-pathologic conferences (CPCs) and in clinical problem-solving exercises (Jain BP. An investigation into method of diagnosis in clinicopathologic conferences (CPCs). Diagnosis 3 (2) https://doi: 10-1515/dx-2015-0034; Jain BP. Why is diagnosis not probabilistic in clinical-pathologic conferences (CPCs): Point. Diagnosis. https://doi.org/10.1515 dx-2016-0012). In one such exercise, for example, acute myocardial infarction is suspected and diagnosed accurately after testing in a healthy 40 year old woman with highly uncharacteristic chest pain in whom its presentation is atypical (Pauker SG et al. How sure is sure enough? N Engl J Med. https://doi.org/10.1056/NEJM199203053261007).

In conclusion, it is not heuristic of representativeness, but lack of awareness of atypical presentations of a disease due to lack of experience, which leads to failure to suspect a disease with an atypical presentation. Therefore, it is not this heuristic, but lack of experience, which is not only a potential, but a real danger in the diagnostic process.

Bimal Jain, M.D
Mass General Brigham/Salem Hospital

Interdisciplinary Systematic Review: A Novel Approach to Evidence Synthesis
:: Sahanika Ratnayake

Abstract
Reporting on the newly started project *Interdisciplinary Systematic Review: A Novel Approach to Evidence Synthesis*

Keywords
intervention; EBM; mechanistic evidence; epistemic injustice; interdisciplinary systematic review; evidence synthesis.

How to Cite
Ratnayake, S. Interdisciplinary Systematic Review: A Novel Approach to Evidence Synthesis. The Reasoner, 19(2). http://doi.org/10.54103/1757-0522/28330

Though ”evidence-based” is now a byword across a number of fields such as public health, medicine and law, the methodological question of how to establish whether a particular intervention or policy brings about the desired effect is somewhat lacking. Traditional evidence reviews, which aggregate and analyse data from individual studies, tend to focus on the question of *whether* an intervention works, rather than *how* or *why* an intervention produces its effects. Such orthodox evidence reviews of the kind described by the Cochrane Institute, tend to rely on limited studies, namely clinical trials, neglecting or devaluing the critical contribution of

http://doi.org/10.54103/1757-0522/28330

mechanistic evidence which provides insight into the latter question.

The exclusion of mechanistic evidence in the current system of evidence review raises both epistemic and ethical issues:

On the epistemic side, the exclusion of mechanistic evidence makes it difficult to gain a comprehensive picture of efficacy and effectiveness. Understanding the mechanism of action can point towards factors that may compromise or enhance the efficacy of the intervention, for instance in the case of drug interactions or genetic variations. Additionally, mechanistic evidence is crucial for translating interventions in clinical settings into real world settings or public policy. Unlike traditional efficacy research, mechanistic research draws on work across a number of disciplines which is vital for implementation and wider adaptation, particularly of complex interventions. For example, in appraising whether a psychological intervention in prisons affects recidivism rates, researchers may draw on not only trials for the psychological intervention, but social science research into the factors that affect recidivism rates and how they interact with the intervention.

Ethically, the neglect of mechanistic evidence can result in replicating patterns of structural injustice and epistemic injustice. For example, in public health, it is widely acknowledged that individuals from BAME backgrounds tend not to seek and have issues accessing mental health treatment or certain types of cancer screening. Research into service and practitioner experiences as part of gaining mechanistic insights can contribute to solving such problems of access and capture the service users experiences of treatment, which will mitigate the possibility of epistemic injustice.

In light of these shortcomings in traditional evidence review, the project will develop an alternative methodology — Interdisciplinary Systematic Review (ISR), designed to integrate mechanistic evidence from diverse disciplines with conventional efficacy research. The aim is to provide a deeper understanding of whether an intervention works by exploring how it brings about this effect. Additionally, in expanding the types of evidence eligible for review, ISR intends to address the ethical issues posed by restricting the evidence base.

In developing ISR, the project will conduct an extensive review of the effectiveness of face masks and mask mandates in controlling respiratory infections. Building on an earlier review, this case study will draw on a variety of study designs from across a range of disciplines such as: physics and engineering (masks are physical barriers with physical, chemical and electrostatic properties), biomedical sciences (the interventions target particular pathogens with particular patterns of spread), psychology (the impact of mask policies and mandates will vary depending on adherence to those mandates and policies), and law and social policy (law and policy are used to prescribe and enforce mask mandates).

Previous systematic reviews on masks have yielded misleading findings, largely due to methodological limitations in the current system of evidence review. For instance, certain reviews have concluded that there is only moderate or poor evidence for masking as a result of downgrading studies for not blinding participants, which is unfeasible in the case of masking. In contrast, a preliminary review of multi-disciplinary evidence for masking suggested that masking mandates had been effective during the pandemic. In addition to exploring whether ISR resolves the ethical and epistemic problems posed by orthodox systematic reviews, the case study will allow ISR to be distinguished from other existing types of review which draw on mechanistic evidence such as realist and narrative reviews.

In addition to delivering formal ISR guidelines, checklists and toolkits, the project will have an extensive dissemination strategy to engage with policymakers, regulatory bodies and service user organisations via relationship building and a series of workshops and training courses.

The project team consists of Jon Williamson (Principal Investigator); Trish Greenhalgh and Rebecca Helm (Co-Investigators); Luana Poliseli and Sahanika Ratnayake (Research Associates). Please get in touch if you would like to hear more or collaborate: sahanika.ratnayake@machester.ac.uk

SAHANIKA RATNAYAKE
University of Manchester

M. D’ Agostino, S. Modgil and C. Larese. *Depth-Bounded Reasoning. Volume I: Classical Propositional Logic. College Publications, London. 2024, xvii + 225, ISBN 978-1-84890-442-2*

:: Alejandro Solares-Rojas

Abstract

D’Agostino et al. recently launched book on the College Publication series on Logic and Bounded Rationality is reviewed. Applications to human-oriented AI are emphasized.

Keywords

Depth-bounded Boolean logics; bounded rationality; book review

How to Cite

Solares-Rojas, A. *M. D’ Agostino, S. Modgil and C. Larese. Depth-Bounded Reasoning. Volume I: Classical Propositional Logic. College Publications, London. 2024.* The Reasoner, 19(2). http://doi.org/10.54103/1757-0522/28422

Can logic-based ’slow thinking’ models complement machine learning ’fast thinking’ methods in meaningful and practical ways? According to Neuro-Symbolic AI, logic-based models constitute a promising interface between opaque machine-oriented methods and human-oriented design and audit, as these models rely on widely developed theories of inference and argumentation.

http://doi.org/10.54103/1757-0522/28422

However, standard theories are highly idealized in that they model omniscient agents who can recognize all logical consequences of their assumptions. Approaches like the one proposed in this book aim to provide realistic models, accounting for the cost that inferences imply to resource-bounded agents. These approaches recognize that making inferences often exceeds cognitive resources and is generally computationally hard, as evidenced by the likely intractability of Classical Propositional Logic (CPL). In particular, the book's approach approximates classical-logic reasoning by defining a hierarchy of increasingly stronger, yet tractable, sub-logics that converge to CPL. These sub-logics can be intuitively associated with resource-bounded agents who approximate ideal omniscient ones. The conceptual basis of the approach is the distinction between actual and virtual information, namely, operational information that is practically available to the agent, versus hypothetical information that the agent does not actually hold but temporarily assumes as if she did.

The Prequel explains this distinction using the sudoku puzzle. Reasoning with actual information corresponds to steps performed using an ink pen, while reasoning with virtual information corresponds to steps requiring a pencil and eraser. A typical reasoning pattern based on actual information is the Single Candidate Principle (SCP): using only available information and known constraints, a single candidate is determined by excluding all other options. By contrast, some reasoning patterns essentially require the introduction of alternate hypotheses and keeping track of their consequences, i.e., the introduction of virtual information that is not even implicitly contained in the information held by the agent. The more nested use of virtual information required, the harder the deduction. The maximum number of these nested uses yields a sensible measure of the difficulty or depth of the deductions. Unfortunately, standard models are structurally inadequate to account for a notion of depth that is semantically well-founded, in the sense that the meaning of the logical operators remains the same throughout the corresponding hierarchy of approximations. Therefore, the approach resorts to non-standard semantics and proof theory. *Chapter 1* focuses on the basic, 0-depth approximation related to easy deduction steps that depend only on understanding the operators' meaning and applying the SCP accordingly. An informational semantics for the operators is given, based on the notions of informational truth and falsity. These notions satisfy a corresponding version of Non-contradiction but not of Bivalence, under penalty of omniscience. CPL's standard semantics is thus not suitable, so two equivalent alternative semantics are explored: constraint-based and non-deterministic. Both fix the meaning of the operators solely in terms of actual information, with no use of virtual information at all, and yield a notion of implicit-information extraction that is easy and 'local'. The 0-depth approximation is Tarskian, has no tautologies, and there is no functionally complete set of operators for it. So, different operators' choices define actually different 0-depth logics. Under any choice, however, the induced logic can be decided in quadratic time. This is shown via a proof-theoretic characterization that is a non-standard Natural Deduction system, where the introduction and elimination rules have a linear format, involve only actual information, and correspond to typical deduction patterns. These rules are taken to fix the meaning of the operators and are indeed sound and complete with respect to any of the two equivalent semantics. Quadratic-time tractability follows from the system's satisfaction of the subformula property, meaning the rules stand not only for easy but also 'local' steps. Moreover, the proof system enjoys an inversion principle, and derivations with the subformula property are uniformly shorter than those without it.

Chapter 2 studies two alternative, albeit not equivalent, ways of characterizing the approximations of greater depth, yielding the weak or strong version, respectively. Both ways are characterized semantically and proof-theoretically, sharing the same 0-depth basis and overall conceptual framework. In both, a deduction's depth is identified with the maximum nested use of a single rule that implements Bivalence and controls the introduction of virtual information. However, the rule format and the specific induced measure of depth distinguish between weak and strong approximations. The strong format can represent non-nested applications within the same derivation, while the weak format does not. Thus, the depth of the 'same' deduction may well be lesser in the former format than in the latter. Their semantics vary according to the format, but they are essentially recursive extensions of the 0-depth approximation semantics. The main point is that depth-increase corresponds to the indispensable introduction of information that cannot otherwise be obtained by the operators' meaning and SCP applications. This intuitively involves more costly reasoning steps whose cost increases proportionally to their nesting. Thereby, hierarchies of approximations are defined, where up to k nested uses of virtual information are allowed, and whose tractability is guaranteed whenever k is fixed and the subset of formulas that can be conclusions of the respective Bivalence-rule or the introduction rules is suitably restricted. Fewer restrictions yield deductively stronger approximations, and their suitability depends on the intended application. Remarkably, the approach provides a logical measure of the difficulty of single deductions, their tractability being a by-product. The k-depth approximations, $k > 0$, may be Tarskian depending on the mentioned subset restrictions. Moreover, tautologies increase with k and derivations can be normalized. Furthermore, non-refutational normal proofs enjoy the non-contamination property, which is a sort of variable-sharing property that bans irrelevant applications of *ex-contradictione quodlibet*.

This last property is useful for applications in argumentation theory, which is the topic of *Chapter 3*. Argumentation provides a unifying and promising setting for a variety of non-monotonic logics. Specifically, it allows for dialogues between agents, where they reason together by exchanging information, resolving conflicts, and finding joint deliberations. However, standard models impose counterintuitive and highly idealized requirements on agents. First, they usually leave implicit the proof-theoretic means by which arguments are constructed and thus also their persuasive force. Second, they imply omniscience by assuming that all arguments defined by a base can be constructed and included in the corresponding framework, and that the legitimacy of each argument is verified by checking, prior to inclusion, that its premises are consistent and non-redundant. These assumptions depart from real-world argumentation and are intractable.

The book's approach facilitates models that are suitable for practical desiderata and rational with respect to resource-bounded agents. Specifically, a notion of argument is given that distinguishes between premises that the agent commits to and those 'supposed for the sake of argument'. This allows for realistic models of premises' inconsistency demonstration via dialectical inter-agent argumentation. Then, intractable checks for arguments' legitimacy are dropped in favor of frameworks that include only the arguments within the agents' construction capabilities. The approach takes normal proofs as explications of arguments, and the resource-boundedness notion is exploited in that agents may still be credited as rational when tractably constructing arguments up to a given *k*-depth. Accordingly, the non-contamination property of *k*-depth normal proofs stops the generation of obviously redundant proofs by tractable means.

Chapter 4 discusses how the approach can help solve philosophical problems arising from the view that CPL is informationally trivial. Orthodoxy holds that, in any valid deduction, the information of the conclusion is implicitly contained in the information of the assumptions. Valid deductions are said to be analytic in the semantic sense that their validity depends merely on the meaning of the operators. However, CPL's probable intractability strongly suggests that the conclusion of certain complex inferences may convey information that is not contained in the assumptions, in the objective sense that there is probably no feasible procedure for extracting it. Therefore, these inferences should be regarded as synthetic. According to the approach, CPL is 'trivial' only for omniscient agents, and not for realistic agents who consume resources when reasoning. Only 0-depth deductions are analytic, whereas deductions of greater depth are increasingly synthetic, in that their validity does not depend solely on the operators' meaning and their conclusion conveys information that is not even implicitly contained in the assumptions.

The *Conclusion* discusses ideas and methods closely related to the approach's non-standard semantics and proof theory. A brief overview of the state-of-the-art of an emerging research program is also given, which spans from covering logics other than CPL to applications in Probability.

In summary, the book provides logic-based models of resource-bounded agents within a robust and well-motivated conceptual framework. These models are useful in a range of practical and multidisciplinary applications, particularly in human-oriented AI, which is of current importance. I would have liked to find more pointers to alternative approaches for designing these models in the book. However, scattered references were perhaps avoided, and a robust survey definitely deserves independent treatment. I believe that the book's contents are of interest and accessible to a wide audience, having a clear multidisciplinary appeal at the intersection of AI, Economics, Philosophy, and Cognitive Science, to name a few. Except for some easily recognizable typos, the book is generally well-written and strikes a balance between technicalities and the intuitions underlying them. The content difficulty is kept to a minimum, requiring basic to medium technical training from the readers. Still, given that the book is well-organized and generally self-contained, readers can easily select content more suited to their background and interests, relying on pointers to more basic or complementary material. The book constitutes an excellent start for the series on *Logic and Bounded Rationality*, and I am sure that it will be a key reference for years to come.

Alejandro Solares-Rojas

Universidad de Buenos Aires