The Reasoner

Volume 17, Number 6
November 2023

thereasoner.org
ISSN 1757-0522

Contents

Features: Focussed issue on *The Handbook of Rationality*

Guest Editorial

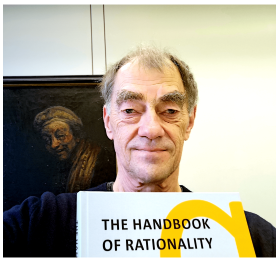
This feature presents *The Handbook of Rationality*, edited by Markus Knauff and Wolfgang Spohn and published in 2021 by MIT Press, Cambridge, MA. We are happy to announce that this Handbook is now *open access* and thus freely available to everyone, thanks to the generous support of MIT. The open-access URL of the Handbook is

https://doi.org/10.7551/mitpress/11252.001.0001

All chapters can be read, downloaded, and printed individually.

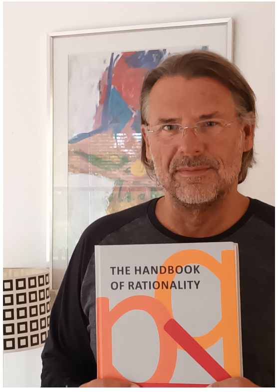
In this feature, a few aspects of the Handbook are highlighted by some of its authors. This contribution and the next by Sangeet Khemlani are mainly about the introductory chapter of the Handbook, briefly explaining a systematic structure in the multitude of accounts of rationality. While we say a few more words about the background of our systematization from the philosophical perspective, Khemlani points out how important this system is for psychological research on human rationality. The three contributions by Hans Rott, David Over, and Gabriele Kern-Isberner emphasize aspects of conditionals from the side of philosophy, psychology, and AI. Finally, Werner Raub succinctly summarizes the use of rational choice theory in the social sciences.

We believe that the Handbook is the most authoritative and comprehensive guide through the field of human rationality as it stands today. Its broad coverage of topics mainly from psychology and philosophy, but also from neuroscience, economics, and the social sciences, is unique in the field and offers researchers and students a valuable resource of cutting-edge knowledge about human rationality.



Written by internationally leading experts, the 65 chapters of the Handbook cover in the main normative and descriptive theories of rationality – how people ought to think, how they actually think, and why they often deviate from what we call rational. It also proposes a novel system for categorizing and evaluating concepts, theories and empirical findings on human rationality from different disciplines. Following the basic distinction between theoretical and practical rationality, the book first considers the theoretical side, including normative and descriptive theories of logical, probabilistic, causal, and defeasible reasoning. It then turns to the practical side, discussing topics such as decision making, bounded rationality, game theory, rational choice theory, deontic and legal reasoning, and the relation between rationality and morality. Finally, it covers topics that arise in both theoretical and practical rationality, including visual and spatial thinking, scientific rationality, how children learn to reason rationally, and the connection between intelligence and rationality.

The main intention of the *introductory chapter*, written by the editors, is to elaborate a systematic order of the very rich research on human rationality along four dimensions. The first dimension is spanned by the distinction between *theoretical* and *practical rationality*. Theoretical rationality deals exclusively with belief and knowledge as well as their synchronic and diachronic principles, i.e., with epistemic matters. Practical rationality, by contrast, deals with the rationality of desires and actions, or volitional attitudes in general. The two parts are asymmetric. While epistemic rationality can be studied without attending to the practical side, practical rationality presupposes theoretical rationality. There can be no rational pursuit of our goals without epistemic rationality. Arguably, epistemic rationality is also distinguished by having a single aim: truth. In any case, without its relation to truth epistemic rationality cannot be substantially understood. By contrast, practical rationality cannot be characterized in such a unique way, since it is at least doubtful whether we can speak of true norms and values which we rationally ought to pursue.



The second dimension is constituted by the distinction between *normative* and *descriptive theories of rationality*. It is important here that normativity is properly understood. Which norms hold in a given group or society is nothing but an empirical, though often difficult, issue. Genuine normativity, by contrast, can be grasped only from the first-person perspective, where I ask: What ought I to do or believe? Note that it is always an open question whether I should follow the norms empirically given (be it by the state or God or whatever). Philosophers intensely discuss norms of rationality in this genuine sense, while psychologists attempt to determine the empirical functioning of rationality. However, the distinction transcends the disciplines; for example, economics is still ambiguous. As clear as it is that both kinds of theories exist, as unclear is their relation. We observe a deep confusion about this. A natural response might be that there can't be any relation at all, since David Hume taught us that no logical inference carries us from "is" to "ought" or from "ought" to "is". However, denying any relation between the normative and the descriptive is not helpful. There is obviously some interaction across the divide that we need to understand. We argue for the claim that there are countless defeasible (not logically cogent) relations between the normative and the descriptive. The origin of these relations lies in the empirical fact that human beings are receptive for norms (even in the genuine sense). Clearly, this receptiveness, our norm compliance, varies enormously. Still, there is at least a weaker or stronger presumption that norms are realized, and reversely, that actual behavior follows some norms. This presumption is very often defeated, but it exists and is responsible for the interaction of the normative and the descriptive.

The third dimension is opened by the distinction between *individual* and *collective* or *social rationality*. Cognitive psychology as well as, for example, traditional epistemology are largely on the individual side. Classically, the distinction is exemplified by individual decision theory and game theory, which is about the interaction of several players. But in the meantime, the distinction is instantiated by many other accounts and investigations represented in the Handbook. Again, the deep question is about the relation between the two sides. For example, epistemic game theory tries to reduce game theory to decision theory. And the principle of methodological individualism generally postulates the reducibility of social macro-laws to micro-principles of rational choice. This question will still occupy us for a long time.

The fourth dimension is not clearly perceived in the literature, we found. It is given by an *output-oriented* versus a *process-oriented perspective on rationality*, as we call it. We can ask for the rationality of the result of a reasoning process. Is that the result called for by standards of rationality? Or we may ask for the rationality of the various steps of a reasoning process. What are the rational rules governing the process? In logical terms, we might say that semantics is output-oriented in defining logical consequence, while calculi are process-oriented in providing rules for implementing logical consequence. Clearly, though, the distinction is very general and not restricted to logics. Again, it is an interesting issue which roles normativity and descriptivity play on both sides of the distinction.

Overall, the four dimensions produce sixteen cells. They are not filled equally, some are even empty. But it is our claim – the entire Handbook is kind of proof of that claim – that every existing attempt in rationality research fits into one or several of those cells. We hope that this schema offers a fruitful systematization for future rationality research. The handbook also describes the cognitive, cortical, and evolutionary preconditions of rationality, and illustrates some historically important paradigms of rationality research and the different intellectual traditions in different disciplines – where they intersect, fall apart, and converge. The introductory chapter is highly recommended to all readers who want to orient themselves in the widely ramified research on human rationality.

Wolfgang Spohn
University of Konstanz, University of Tübingen
Markus Knauff
University of Giessen

Finding your way through *The Handbook of Rationality*

A common approach to reading volumes such as *The Handbook of Rationality* is summarized by this algorithm:

1. Find and read a chapter most relevant to a personal interest,
2. skim some of the other chapters it cites,
3. exit the volume,
4. discover a new interest,
5. return to Step 1.

The approach is sensible, particularly for reference collections as heterogeneous as an overview of recent analyses of rational thought. But if you follow this algorithm with *The Handbook of Rationality*, you'll neglect an important insight: philosophers and psychologists exhibit certain patterns in the ways they construe rationality.

The study of rational thought is vast: its history stretches to antiquity and its contemporary analyses appear to encompass almost every scholarly discipline. And so, Knauff and Spohn intended the Handbook to serve as a compendium for the way rationality manifests both in the study of human thinking and in the practical exigencies of daily life. Rationality, after all, has value: to establish that somebody erred in their thinking is to permit the possibility of correction. To establish that they made a rational decision is to hold their pattern of thinking up as an example for others to follow.

The Handbook therefore includes dozens of perspectives on what it means to be a rational individual in the real world. It contains contributions from philosophers, cognitive scientists, decision scientists, and theorists who wrote on topics such as

- rationality from an evolutionary standpoint,
- rationality as it emerges from mental simulation,
- Bayesian analyses of rationality and the dynamics of subjective probabilities,
- social, communicative, and adaptive rationality,
- rational ways of revising beliefs,
- rationality in moral and legal thinking.

But the inclusion of so many different perspectives came at a cost. There exists scant consensus about what constitutes rational behavior from one pattern of thinking to the next: rational legal thinking may seem qualitatively distinct from rational spatiotemporal reasoning. So, readers who digest one chapter of the Handbook at a time may be tempted to conclude that no pattern exists which links the different perspectives on what it means to be rational.

Knauff and Spohn urge against this temptation: they begin the Handbook by offering readers holistic ways to think about where the different approaches to rationality intersect and where they conflict. They focus their introductory chapter on four separate dichotomies that help to categorize theoretical treatments together. Perhaps one that may be familiar to some readers is the distinction between *normative* and *descriptive rationality*: cognitive scientists often treat branches of mathematical philosophy, such as classical logic, probability theory, and decision theory, as describing the sorts of inferences competent reasoners should make given infinite resources and infinite time (of course, there exist normative accounts beyond these three frameworks). These theories are normative in that they establish idealized responses to idealized scenarios. In contrast, descriptive rationality concerns the development of accounts of how and why reasoners make everyday inferences, and how they repair them. Psychological investigations of the algorithms by which people reason fall into this category. They can appeal to sets of interoperating systems or mechanisms, such as "system 1" (intuitive) versus "system 2" (deliberative) thinking popularized by Kahneman (2011: *Thinking, Fast and Slow*, Farrar, Straus and Giroux) and Stanovich and West (2000: Advancing the rationality debate, *Behavioral and Brain Sciences* 23, 701–26), one of which is corrective and slower in nature and the other of which operates rapidly. Such descriptive accounts of rationality help explain why reasoners often deviate from a norm: they may have followed a useful heuristic, be prone to an alluring bias, or they may substitute information they don't have or are uncertain about for information that's available or salient.

But as Knauff and Spohn observe, theorists who support some normative theory or else some descriptive theory often argue against one another without acknowledging where their views coincide. They point out another useful dichotomy: one between *process-oriented* and *output-oriented perspectives on rationality*. The dichotomy mirrors one that David Marr (1982: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman) proposed between the algorithmic and computational levels of explanation.

Process-oriented rationality focuses on how individual reasoners achieve rational thought: the mental representations they build, the algorithms they carry out, the errors they make in the process. Output-oriented rationality concerns how beliefs should change given external influences, such as the acquisition of some new piece of information. It's possible to build descriptive accounts of rationality that focus either on rational processes or on rational outcomes. It may even be the case – as Knauff and Gazzo Castañeda (2023: How are beliefs represented in the mind?, *Thinking & Reasoning* 29(3), 416–26) recently observed – that many disagreements can be curtailed by appropriately classifying theoretical disagreements based on whether they concern processes or outcomes.

The editors highlight two additional and intuitive dichotomies to classify treatments of rationality. One is between *theoretical* and *practical rationality*: you may, for instance, agree that in theory, it is most rational to create a tax on carbon to curb climate change, but that, practically speaking, it is irrational to advocate for any policy that would hurt your personal finances, and a carbon tax could do precisely that. Theoretical rationality focuses on rational justifications of beliefs, positions, and explanations, whereas practical rationality is the rationality of actions and consequences. Likewise, some treatments of rational thinking focus on what it means for an individual to process information optimally, whereas others focus on how social groups maintain beliefs, or how individuals take their social context into account when making rational decisions. Hence, accounts of *individual* and *social rationality* may look quite different from one another.

These four dichotomies: normative versus descriptive rationality, process- versus outcome-oriented rationality, theoretical versus practical rationality, and individual versus social rationality, can place every chapter of the Handbook into its appropriate context. These dichotomies don't just help bring order to disparate sets of analyses: they may also help in the development of new theoretical models of rational thinking. And they may be useful in classifying everyday patterns of thinking and reasoning. For instance, people in Western cultures appear to exhibit antipathy towards euthanasia (see Goodwin, under review), despite the fact that there exist scenarios for which euthanasia may be the only ethical option amongst a set of terrible alternatives. Indeed, active euthanasia is illegal throughout the United States and many other nations. Is this antipathy rational? From the perspective of social rationality, perhaps so: permitting any form of killing an individual may be exploited by other members of a society. From an individual perspective, perhaps not: individuals in extraordinary pain with only a few days to live may consider it profoundly irrational for a country to mandate living. Hence, it may be facile to describe the policy and the law as rational or irrational: what matters is how it's construed and by whom.

In sum, Knauff and Spohn's *Handbook of Rationality* surveys a dense, diverse, and controversial landscape of rational thinking. Heated debates abound, but the editors show how to navigate them. Before you read any element of the Handbook – including my own review of the psychology of syllogisms – I urge you to read Knauff and Spohn's bracing introductory chapter.

Sunny Khemlani
US Naval Research Laboratory

"If the weather is nice, we'll go hiking" – Supposition and difference-making

It is hard to imagine what treasure of philosophical and mathematical insights Frank Ramsey would have unearthed had his life not been so terribly short. He died in 1930 at the age of 26. But his name is mentioned many times in *The Handbook of Rationality* edited by Markus Knauff and Wolfgang Spohn. Most of the time he is referred to as one of the originators of the subjectivist theory of probability and of modern decision theory – pillars of important theories of rationality. The topic of my contribution, however, relates to a single footnote in Ramsey's paper "General Propositions and Causality" (written in 1929), which formulated what has become known as "the Ramsey test" for conditionals. It ties conditionals to suppositional reasoning and says that a conditional *If A then C* is evaluated by first supposing that *A* and then checking, on this supposition, the acceptability or probability of *C*.

The Ramsey test is referred to in nine out of the 65 articles in the Handbook (it is quoted on p. 296; all page references will be to the Handbook). It has caught the interest of psychologists at least as much as the interest of philosophers and logicians (though Ramsey himself was not a psychologist). Indeed, the Ramsey test is associated with nothing less than a paradigm change in the psychology of thinking that happened about two decades ago – essentially a change from classical logic to probability theory as the basic framework for the study of good reasoning. In the Handbook, Ramsey's informal footnote is interpreted as yielding "the Equation" that identifies the probability of a conditional with the conditional probability of its consequent given its antecedent, $\Pr(\textit{If } A \textit{ then } C) = \Pr(C \mid A)$. It should be noted, though, that the Ramsey test formulates an attractive idea not only for probabilists, but also for researchers studying propositional attitudes in a qualitative way (Nute and Cross 2001: Conditional Logic, in Gabbay and Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edn., vol. 4, Kluwer, 1–98; Rott 2017: Preservation and postulation – Lessons from the new debate on the Ramsey test, *Mind* 126, 609–26).

The "new paradigm psychology of reasoning" did two things, one more general than the other: first, it replaced logic by probability theory, and second, it replaced the interpretation of conditionals through the truth-functional material conditional by the interpretation through conditional probabilities. It was certainly right in dismissing the idea that conditionals in natural language are material conditionals. I doubt that any logician has ever advocated this idea (I don't count Grice as a logician), and the material conditional interpretation is obviously wrong for counterfactuals which use the same *if … then* connective as indicatives. But the paradigm shift sold "binary" deductive logic at less than fair value. Logic, even bivalent logic, is not committed to identifying natural language conditionals with material conditionals: there are modal logics, conditional logics, relevance logics, connexive logics and more, which all conceptualize conditionals as being *very* different from material conditionals. Perhaps the pioneers of the new paradigm missed these developments and threw out the baby with the bathwater.

Be that as it may, they provided many noteworthy arguments and substantial empirical support for the thesis that the probability of a conditional *If A then C* does not equal the probability of the material conditional, $\Pr(A \supset C) = a + c + d$, but the conditional probability $\Pr(C \mid A) = a/(a + b)$, which is normally lower than the former (see table 1). In this view, the (probabilities of the) $\neg A$-worlds don't matter at all for the evaluation of the conditional.

	C	$\neg C$
A	a	b
$\neg A$	c	d

Table 1: The probabilities of the four possible combinations $A \wedge C$, $A \wedge \neg C$, $\neg A \wedge C$ and $\neg A \wedge \neg C$ sum up to 1: $a + b + c + d = 1$.

The suppositional interpretation, however, has an unwelcome feature: for a conditional *If A then C* to be true or accepted, it suffices that *C* is true or accepted and *A* does not interfere with *C*'s truth or acceptance. *A* does not have to be positively linked or connected to *C*, it is enough if it does not get in the way of *C*. But this is not how conditionals are used in natural-language conversations. Consider the conditional in the title of this contribution. In ordinary circumstances, it indicates that good weather *supports* our plan for a hike. There are many other locutions that make a similar point: good weather is (causally or evidentially) *relevant* to our hike, it is a *reason* for, or *makes a difference* to, our hiking, our hike *depends* on good weather, one can *infer* from good weather that we'll be going for a hike. Similar messages are conveyed by most conditionals used in natural language. On the other hand, saying "If the weather is bad, we'll go hiking" would be misleading in ordinary circumstances, unless one adds "anyway" at the end of the sentence or replaces "if" by its concessive cousin "even if". Some conditionals are not meant to indicate positive relevance, such as the conditionals used in mathematics and in legal texts (where the material conditional reading seems appropriate) and perhaps also the conditionals presented in psychological tests (which often have an artificial content). But it seems safe to say that most of the time, typically, the conditionals uttered in ordinary conversations indicate that the antecedent makes a difference for the consequent. If this is not part of the meaning of such conditionals, it is a conventional implicature; if it is not part of the semantics, it is a pervasive feature of the pragmatics of conditionals.

This idea doubles the Ramsey test, as it were, and contrasts the supposition of good weather with that of bad weather. Thus the antecedent is not only contextually sufficient, but also, in some way, contextually necessary. "Conditional perfection" is satisfied, but this does not turn a conditional into a biconditional. Similar ideas are followed within the inferentialist camp, which posits, given certain background beliefs, a compelling argument from A to C (but not from the background beliefs alone). As Igor Douven, Shira Elqayam, Niels Skovgaard-Olsen and their colleagues have shown, such an interpretation has considerable experimental support (pp. 299–300, 399–400; there is also evidence against it, the case is still open). Inferentialists are often wary of defining the inferential connection between antecedent and consequent by logical or probabilistic means. But the idea that antecedents are difference-makers can be captured by formal accounts. One can either use the Ramsey test for identifying a qualitative contrast regarding C between the suppositions that A and that $\neg A$. Or one might want to bring in measures of "evidential relevance" (which are beautifully reviewed by the late Arthur Merin in chapter 4.3 of the Handbook), or alternatively the ΔP measure, which is defined as $\Pr(C\,|\,A) - \Pr(C\,|\,\neg A) = a/(a+b) - c/(c+d)$ (provided that A is positively relevant for C, pp. 273, 441–2). Obviously, this latter measure gives values lower than the conditional probability $\Pr(C\,|\,A)$. And it makes the (probabilities of the) $\neg A$-worlds matter again. Difference-making conditionals share this latter feature with material conditionals – but in a very different way.

Hans Rott
University of Regensburg

Independence and rationality

The Handbook of Rationality covers an impressive variety of topics, perspectives, approaches, and disciplines in the inexhaustible study of human rationality. My research field, the psychology of reasoning, is very well represented in the Handbook, but there is nothing in it on how people reason and convey information about causal (or other) independence, e.g., using mental models (Johnson-Laird, chapter 2.3) or Bayes nets (Hartmann, chapter 4.2; Waldmann, chapter 7.2).

This information is clearly sometimes necessary for rationality, and it would appear that we often use conditionals to convey it. Suppose some parents whose children do not have autism are worried about the MMR vaccine. A doctor could reassure them by asserting:

(1) If your children are vaccinated, they will not get autism.

To be clear as possible about what is being conveyed by the assertion of (1) in a given context, the doctor might add:

(2) If your children are vaccinated, they will not get autism, and if they are not vaccinated, they will not get autism.

The doctor would be conveying, in this context, that developing autism is independent of being vaccinated. Let us say that a conditional used, in a given context, to convey information about independence is an *independence conditional* (Cruz and Over 2023: Independence conditionals, in S. Kaufmann, D. E. Over, and G. Sharma (eds.), *Conditionals: Logic, Linguistics and Psychology*, 223–33). It is possible to use (1) as a *dependence conditional*, making it the claim that the vaccine was supposedly one for preventing autism. Asserting additionally (2) would clarify that (1) was used as an independence conditional.

Some independence conditionals are also *concessives*. A concessive conditional has *even if* in it instead of *if*. The former could replace the latter in this independence conditional about an easy test:

(3) If you do not revise, you will pass that test.

A concessive conditional, *even if not-p, q*, is often used when $p\ \&\ q$ holds in that context, and an independence conditional *if p then q* validly follows by an *and-to-if*, or *centering*, inference, according to a number of accounts of conditionals, including possible-worlds and interventionalist analyses (Starr, chapter 6.1; Pearl, chapter 7.1), and the probabilistic theories that imply *the Equation*. This states that the probability of the conditional is the conditional probability, $P(\textit{if } p \textit{ then } q) = P(q\,|\,p)$, which implies in turn that it is probabilistically valid, *p-valid* (Over and Cruz, 6.2), to infer *if p then q* from $p\ \&\ q$, because $P(p\ \&\ q) = P(p)P(q\,|\,p) \leq P(q\,|\,p)$. Both *if p then q* and *if not-p then q* can be used as independence conditionals when $P(q\,|\,p) = P(q\,|\,\text{not-}p)$.

Several chapters in the Handbook refer to the Equation (Evans, chapter 1.2; Chater and Oaksford, chapter 4.5; Gazzo Castañeda and Knauff, chapter 5.4; Oberauer and Pessach, chapter 4.6; Over and Cruz, chapter 6.2). It is highly confirmed in many experiments, and some studies have also confirmed the validity of centering (Oberauer and Pessach, chapter 4.6). But the Equation and centering have not been supported for *missing-link conditionals* (Oberauer and Pessach, chapter 4.6):

(4) If Peter is wearing a blue shirt, the sea levels will rise.

By our definition, (4) is an independence conditional, but it is pragmatically problematic, unlike (1) and (3), which could also be said to be "missing-link" conditionals. We need a technical term specifically for conditionals like (4), to distinguish them from pragmatically acceptable independence conditionals such as (1) and (3). Let us call them *Walrus conditionals* after the Walrus in Lewis Carroll's poem, "The Walrus and the carpenter", who says that it is time talk about pragmatically unrelated things: shoes, ships, sealing-wax, cabbages, and kings.

The experiments in the literature which do not support the Equation and centering use Walrus conditionals. In the first studies of this general type, Skovgaard-Olsen, Singmann, and Klauer (2016: The relevance effect and conditionals, *Cognition* 150, 26–36) compared conditionals like the following:

(5) If Mark presses the power button on his TV, then the TV will be turned on.

(6) If Mark is wearing socks, then his TV will be malfunctioning.

In the above, (5) is a pragmatically acceptable dependence conditional, and (6) is a pragmatically unacceptable independence conditional, i.e., a Walrus conditional. This confound could be avoided by comparing pragmatically acceptable dependence conditionals with pragmatically acceptable independence conditionals, like the following (easy to imagine as about a faulty TV):

(7) If Mark presses the power button on his TV, then the screen will remain blank.

Still, it was important for Skovgaard-Olsen et al. (2016) to establish experimentally that there is a problem with Walrus conditionals which needs to be explained. One explanation is (truth condition) *inferentialism*: the hypothesis that *if p then q* is *true* if and only if there is a semantic link, like a deductive or inductive one, between p and q (see Oberauer and Pessach, chapter 4.6, for references). This implies the problem with Walrus conditionals is that they are obviously false, but what about acceptable independence conditionals, like (1), (3), and (7)? They must be given a non-inferentialist semantics, which could then imply that Walrus conditionals can be true and so subvert inferentialism (see Cruz and Over 2023 for more on inferentialism). Other possible explanations should be considered.

Skovgaard-Olsen et al. (2016) themselves propose that there is only a default expectation that p gives a reason for q when *if p then q* is asserted (see also Kern-Isberner, Skovgaard-Olsen, and Spohn, chapter 5.3, on Skovgaard-Olsen et al. 2016). Lassiter (2023: Decomposing relevance in conditionals, *Mind & Language* 38(3), 644–68) develops a pragmatic explanation of why Walrus conditionals are unacceptable (see also Bourlier, Jacquet, Lassiter, and Baratgin 2023: Coherence, not conditional meaning, accounts for the relevance effect, https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1150550/full, *Frontiers in Psychology* 14). From this point of view, people's reaction to Walrus conditionals is a pragmatic bias.

In any case, the psychology of reasoning needs to theorize more deeply, and experiment more thoroughly, building on the theories and results described in the Handbook, to explain fully how people reason about independence, and to assess how far they are rational in this respect.

David Over
Durham University

Plausibility and ranking theory for modelling commonsense reasoning

Reasoning is an eminently rational activity and as such one main focus of the comprehensive *Handbook of Rationality*. It has become clear by now that reasoning is largely reasoning about uncertainty. In fact, probability is the major approach to model uncertainty in everyday life, mainly due to two reasons: first, probabilities allow for assigning quantitative degrees of uncertainty to elementary events or possible worlds, respectively, according to a clear axiomatic semantics, and second, conditional statements "if A then B with probability x" can be given an intuitive interpretation via conditional probabilities. It is particularly this second argument that provides significant advantages for representing knowledge and beliefs (and for reasoning with them) beyond classical (propositional or first-order) logic. However, in between these two well-established frameworks of probabilities and logic, there are formalisms expressing possibilities, plausibilities and the like that are based on classical logic while showing characteristic features of probabilities, but in a less restrictive respectively more abstract way. Spohn's ranking theory (on which *The Laws of Belief: Ranking Theory and Its Philosophical Applications* by Wolfgang Spohn, Oxford University Press 2012, is a most comprehensive work) is one of these formalisms that associates possible worlds w with a degree of implausibility $\kappa(w)$ in natural numbers. Worlds with rank 0 are least implausible, i.e., most plausible, and in general, the lower the rank of a world is the more plausible it is. As kind of a normalization condition, there must always be worlds with rank 0, and some worlds may be deemed so implausible respectively impossible that they are assigned the rank ∞. The rank of a proposition A is the rank of its most plausible models. So, a ranking function κ satisfies the *law of disjunction* $\kappa(A \vee B) = \min\{\kappa(A), \kappa(B)\}$. Since natural numbers are equipped with a basic arithmetics, ranks can also be defined for conditionals "if A then B", in symbols $(B \mid A)$, by setting $\kappa(B \mid A) = \kappa(A \wedge B) - \kappa(A)$. This can be rewritten in the form $\kappa(A \wedge B) = \kappa(A) + \kappa(B|A)$, what Spohn called the *law of conjunction*. Moreover, conditionals can be given an intuitive semantics by saying that a ranking function accepts $(B|A)$ if $\kappa(\neg B \mid A) > 0$, or equivalently, if $\kappa(A \wedge B) < \kappa(A \wedge \neg B)$, i.e., if the verification $A \wedge B$ of the conditional $(B \mid A)$ is more plausible than its falsification $A \wedge \neg B$. This meets the human understanding of conditionals quite well; we accept the conditional "birds usually fly" (in symbols: (fly | birds)) because birds with the ability to fly are more plausible than birds that cannot fly, where nevertheless exceptions are still possible.

Most importantly, this compatibility with nonclassical interpretations of conditionals make ranking functions a convenient common basic tool for nonmonotonic reasoning and belief revision. Belief revision proposes rationality postulates and constructive approaches for revising a belief state by new information. Nonmonotonic reasoning also deals with belief dynamics in that conclusions may be given up when new information arrives (so, the consequence relation is not monotonic, as in classical logic). Both fields emerged in the 1980's (partly) as a reaction to failures caused by classical logic to handle problems in everyday life that intelligent systems like robots were expected to tackle. Knowledge or belief about the world is usually uncertain, and the world is always changing. Therefore, AI systems built upon classical logics failed. So-called preferential models provide an important semantics for nonmonotonic logics. Their basic idea is to order worlds according to normality and to focus on the minimal ones, i.e., the most plausible ones, for reasoning. Likewise, the fundamental AGM belief revision theory (published first in C. E. Alchourrón, P. Gärdenfors, and D. Makinson 1985: On the logic of theory change: Partial meet contraction and revision functions, *Journal of Symbolic Logic* 50(2), 510–30) needs orderings of worlds to become effective. For both fields, ranking functions offer quite a perfect technical tool that also complies nicely with the intuitions behind the techniques. Judea Pearl was probably the first renowned AI scientist to make use of ranking functions; his famous *system Z* (cf. Judea Pearl 1990: System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning, *Proceedings TARK'90*, 121–35) is based on them. Pearl has steadily emphasized the commonsense-related structural qualities of probabilities and interpreted ranking functions as an interesting qualitative counterpart to probabilities. To date, system Z is one of the best and most convenient approaches to implement high-quality nonmonotonic reasoning.

Consequently, ranking functions are deeply connected with nonmonotonic and uncertain reasoning and with belief change, which are core topics in the field of knowledge representation and reasoning. Many researchers make use of them in one way or another even if they rely on more general frameworks. As a prominent example, in "On the logic of iterated belief revision" (1997: *Artificial Intelligence* 89, 1–29), Darwiche and Pearl presented general postulates for the iterated revision of general epistemic states, but illustrated their account with ranking functions. Indeed, ranking functions are particularly well suited for iterated belief change because they can easily be modified in accordance with AGM theory, returning new ranking functions which are readily available for a subsequent change operation. The main AGM operations are revision (adopting a belief) and contraction (giving up a belief), related by Levi and Harper identities. In ranking theory, the connections between these operations are even deeper, since (iterated) contraction is just a special kind of (iterated) ranking conditionalization. Also the results of Kern-Isberner, Bock, Sauerwald, and Beierle in "Iterated contraction of propositions and conditionals under the principle of conditional preservation" (2017: in *Proceedings GCAI 2017*, 78–92) show that iterated revision and contraction can be performed by a common methodology.

Continuing on that, and beyond the practicality and diversity of ranking functions, it is crucial to understand that they are not just a pragmatically good choice but indeed allow for deep theoretical foundations of approaches to reasoning. It is the ease and naturalness with which they can handle conditionals – very similar as probabilities do – that make them an excellent formal tool for modeling reasoning. Given that conditionals are, on the one hand, crucial entities for nonmonotonic and commonsense reasoning and belief change, and, on the other hand, formal entities fully accessible to conditional logics, this characteristics provides a key feature for logic-based approaches connecting nonmonotonic logics and belief change theories with commonsense and general human reasoning. More precisely, conditional ranks give meaning to differences between degrees in belief when observing A versus $A \wedge B$ (see the law of conjunction above), and it is easily possible to preserve these differences under change when using ranking functions. This property has been elaborated as a *principle of conditional preservation* in "A thorough axiomatization of a principle of conditional preservation in belief revision" (Kern-Isberner 2004: *Annals of Mathematics and Artificial Intelligence* 40(1–2), 127–64) giving rise to defining c-representations and c-revisions (where the "c" refers to their common "conditional" base). Interestingly, c-representations emerge from c-revisions when starting from a uniform ranking function, allowing for inductive reasoning from conditional belief bases, similar to Pearl's system Z. Ranking theory is one of the few formal frameworks (probability theory is another) that is rich and expressive enough to allow such a precise formalization of conditional preservation which supports both belief change and inductive reasoning as a common, elegant methodology.

Gabriele Kern-Isberner
Dortmund University

More than just micro-level assumptions: Rational choice theory as a tool for social science

The Handbook of Rationality offers a comprehensive and thorough treatment, with respect to breadth and depth, of theoretical as well as practical rationality, from the normative as well as the descriptive point of view. The Handbook represents perspectives from philosophy, psychology, and the social sciences. Here, I briefly address rational choice theory from a social science perspective (for further discussion see my chapter 10.4, "Rational choice theory in the social sciences", in the Handbook).

Rational choice theory, broadly conceived, is a tool for social science theory construction and explanation, in economics as well as, importantly, disciplines such as sociology, political science, and history. Employing rational choice theory in social science is in line with methodological individualism, accounting not only for the behavior of individual actors but also, and particularly, for phenomena and processes at the level of social systems made up by those actors. Therefore, theory construction and explanation are concerned with two levels, namely, the "micro-level" of actors and the "macro-level" of the respective system.

James Coleman (1990: *Foundations of Social Theory*, Belknap Press of Harvard University Press) has suggested a diagram (Figure 1) that visualizes this approach. Nodes A and D represent propositions describing macro-conditions and, respectively, macro-outcomes. Arrow 4 represents propositions about empirical regularities at the macro-level, say, a statistical association between macro-conditions and macro-outcomes. For example, in theories and models of collective good production (M. Olson 1971: *The Logic of Collective Action*, 2nd ed., Harvard University Press, is a meanwhile classic contribution), the production of the collective good or, respectively, production failure, would be the macro-outcome, while the size of the group of actors that can contribute to and benefit from the production of the good would be one of the macro-conditions. Macro-outcomes and macro-level empirical regularities are explananda at the macro-level. Node B represents propositions describing micro-conditions. Loosely speaking, these propositions refer to "independent variables" in assumptions about regularities of actors' behavior. Arrow 1 represents "bridge assumptions" (R. Wippler and S. Lindenberg 1987: Collective phenomena and rational choice, in J. C. Alexander, B. Giesen, R. Münch, and N. J. Smelser (eds.), *The Micro–Macro Link*, University of California Press, 135–52) on how macro-conditions affect these variables. For example, group

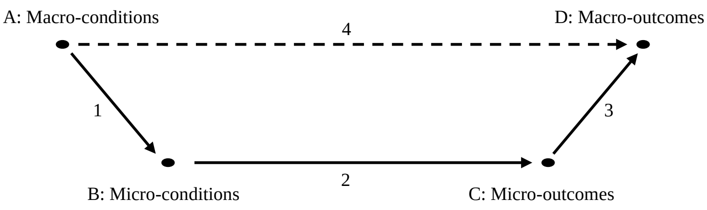


Figure 1: Coleman's micro–macro diagram.

size can shape individual incentives to contribute to collective good production. Node C represents micro-outcomes and the explanandum at the micro-level, namely, descriptions of actors' behavior. In our example, this would be the actors' individual contribution levels. Arrow 2, then, represents a micro-theory, that is, assumptions specifying regularities of the behavior of individual actors. Finally, Arrow 3 represents assumptions on how actors' behavior generates macro-outcomes. We use "transformation rules" (Wippler and Lindenberg 1987) as a label for such assumptions on micro-to-macro relations. As the diagram suggests, the explanandum at the micro-level, descriptions of individual behavior, follows from an explanans comprising a theory of behavior and relevant micro-conditions according to that theory (Arrow 2, Node B), macro-conditions (Node A), and bridge assumptions (Arrow 1). The explananda at the macro-level, namely, descriptions of macro-outcomes (Node D) and macro-regularities (Arrow 4), follow from an explanans comprising a theory of behavior and relevant micro-conditions according to that theory (Arrow 2, Node B), macro-conditions (Node A), bridge assumptions (Arrow 1), and transformation rules (Arrow 3).

The straightforward "place" of rational choice theory in Coleman's diagram is Arrow 2, consistent with interpreting rational choice theory as a "descriptive" (rather than "normative") theory of individual behavior. For example, in game-theoretic models of collective good production, assuming Nash equilibrium behavior or a "refined" equilibrium concept in conjunction with further assumptions as represented by the relevant nodes and arrows of the diagram, one could derive implications on individual contribution levels (Node C) as well as the macro-level of collective good production (Node D), and also on the macro-level association between group size and collective good production (Arrow 4).

However, complementing this straightforward interpretation of rational choice theory as a tool, there are important but less often acknowledged further ways in which the theory is useful for social science. To see this, note that a key issue for theories and explanations in line with an approach visualized by Figure 1 is the careful specification of bridge assumptions and transformation rules that link macro- and micro-levels of analysis. Coleman in particular has argued that social science is often deficient with respect to such specification and there is a meanwhile sizeable literature on this issue (for further discussion and references, see W. Raub, N. D. de Graaf, and K. Gërxhani 2022: Rigorous sociology, in K. Gërxhani, N. D. de Graaf, and W. Raub (eds.), *Handbook of Sociological Science: Contributions to Rigorous Sociology*, Edward Elgar, 2–19).

Game theory is one branch of rational choice theory, focusing on decision making of strategically interdependent actors. Game-theoretic models – more precisely, models of noncooperative games – require the exact specification of the actors' decision situation in terms of the normal form or the extensive form of a game. In light of the diagram in Figure 1, a key contribution of specifying a normal or an extensive form of a game, and a contribution that is not often noticed explicitly, consists precisely in yielding bridging assumptions and transformation rules linking macro- and micro-levels. After all, the normal as well as the extensive form of a game imply how macro-conditions, together with actors' behavior, affect each actor's (expected) payoffs, thus yielding bridge assumptions. Likewise, the normal as well as the extensive form of a game imply the macro-outcomes of actors' behavior, thus also yielding transformation rules.

Consider once again the production of collective goods. Game-theoretic models of collective good production include n-person versions of the Prisoner's Dilemma, the Public Goods Game, and the Volunteer's Dilemma (for brief overview, see W. Raub, V. Buskens, and R. Corten 2015: Social dilemmas and cooperation, in N. Braun and N. J. Saam (eds.), *Handbuch Modellbildung und Simulation in den Sozialwissenschaften*, Springer, 597–626). The normal forms of these games give the number of actors, the strategies for each actor, and each actor's payoff function. Hence, each of these normal forms implies a bridge assumption, namely, how group size, a macro-condition, affects each actor's (expected) payoff. Also, since actors' behavior implies (the amount or likelihood of) collective good production, the normal form implies a transformation rule.

More generally, the specification of rational choice models in social science involves providing clear assumptions on how macro-conditions relate to feasible alternatives between which actors can choose, incentives associated with these alternatives, and actors' information. Also, such specification involves providing assumptions on how actors' behavior relates to macro-outcomes. These assumptions on macro-to-micro as well as micro-to-macro links are key ingredients of social science theory and explanation and they complement micro-level assumptions concerning behavioral regularities such as (expected) utility maximization, game-theoretic equilibrium behavior or variants of such micro-level assumptions that have been developed, for example, in theories of boundedly rational behavior or in behavioral game theory.

Werner Raub
Utrecht University

Dissemination corner

BRIO: From topology to a logic of uncertainty

In the early days of 1841 Ava Lovelace writes: "Mathematical Science shows what is. It is the language of the unseen relations between things. But to use & apply that language we must be able fully to appreciate, to feel, to seize, the unseen, the unconscious." A couple of years later she would go and prepare one of the first computer science papers in history, pioneering the idea of a programmable machine capable of carrying out a number of different tasks. Today, almost two centuries later, we live with the computers that she imagined, and the amount of data and calculations that they perform is so great that no human being can "feel" or "seize". The aim of the BRIO (Bias, Risk and Opacity in AI) project is to use all the tools that reason has to offer to uncover possibly sensitive unseen relations, to try and feel and seize them.

Relations are of course of main importance to science – consider, for example, how relevant equations are to physics – and more so they are to mathematical reasoning. There are many theories of relations, and each gives a perspective on our problem, but today we will discuss one in particular, called *category theory*, that was born around the 1940s in the context of algebraic topology, and out of a very simple "accidental" observation: "many properties of mathematical systems can be unified and simplified by a presentation with diagrams of arrows." Saunders Mac Lane (1978: Categories for the working mathematician)

Consider for example groups – simply put, sets with an operation with identities and inverses, such as the integers with + or the set of permutations of $\{a, b, c\}$ and their compositions – and sets. Most of the time, the relations we are interested in when considering groups are called "group homomorphisms" – that is, functions of the underling sets preserving the operations –, while when considering sets they are simply functions. A *category* is a collection of objects (sets, groups) and their relations or "arrows" (functions, group homomorphisms) satisfying some axioms, so that one has a category of groups, and a category of sets. In this sense, group homomorphisms are to groups what functions are to sets.

Considering statements of this kind not only allows for an organized treatment of structures, but it also allows for the expression of new statements pertaining different entities and relations, putting them in a relation with one another themselves. Using our example categories, one can map each group to its underlying set and each group homomorphism to its underlying function; not only that, one can map each set to the "free group" on the set, meaning the group of words in the alphabet of the set, and each function to the group homomorphism substituting words letter by letter. These two processes are examples of *functors*, or arrows of categories. Extending the paradigm from our previous paragraph: in some sense, functors are to categories what group homomorphisms are to groups and what functions are to sets.

We hope the reader sees that this simple process can be readily developed (what about relations between functors?) and extended to the most various objects (coming from physics, topology, computer science, algebra, chemistry). Not only that, it allows us to establish new relations between different objects, possibly coming from different worlds, possibly *unseen*.

Of course the process of looking for patterns and to relate them, the process of setting a footprint and finding its instances in a multitude of places, is very much familiar to the logician. In fact, categories are well suited to express paradigms of the syntax-semantics kind: in the 1970s Bill Lawvere laid the foundations of a logic based on categories, in which one would describe a category (remember, of objects and relations) acting as the syntax of a certain theory, and the process of taking its (set-based) models as a functor from said category into the category of sets.

This means that we got ourselves some mathematical objects describing not only logic, but *what we do* with it. Of course there is no reason as to why this should only be possible with classical, two-valued logic, and in fact we can use the level of abstraction categories provide to proceed by analogy and extend our now-no-longer-unseen relations to more fancy contexts: modal logic, fuzzy logic, linear logic, and so on.

With the purpose of being able to formulate problems of risk, opacity and bias in AI, we now have a playing field which we can try to combine tools from different theories in. In his piece in The Reasoner Vol. 17 Num. 4, F. A. Genco argues that two that might be of help in this pursuit are that of probability, so that one can try to reign in non-transparent processes, and that of type theory, as in the language describing the computations that said processes undergo.

At the heart of the relation that we wish to express between the two, there is the notion of *dependence*: in the case of probability theory, we have conditional dependence of (random) variables, and in that of type theory, computational dependence of variables. Conditional dependence is intimately connected to the notion of *in-dependence*, and it is meant to describe the process of assessing some probability based on "previous" knowledge, for example the probability of a dice rolling a 6, provided that we know that the outcome is an even number. Type dependency, instead, is a key concept both in mathematical foundations and as a computational model, and it aims to describe entities that might depend on some parameters, as it is frequent in programming languages, for example when defining a list one ought to describe what entries such a list admits, so that `list(int)` and `list(bool)` are, respectively, the type of lists of integers and that of lists of Booleans, hence intrinsically different. In this case we can say that `list` is a type depending on the type of all types.

Of course, conventionally having the same name is no proof of compatibility, and in fact there are many technical differences between the two concepts, starting from their compositionality. Still, some work in this direction as already been done (see, for example, the TPTND calculus), and the question remains: how far can we take this analogy? Can we describe a new "unseen relation", and with it uncover unwanted behaviours of opaque algorithms? Our hope is that using categories we can transport the structure of conditional dependence to that regulating the mutual behaviour of dependent types, so that one can smoothly impress the pattern of the first to the second.

Greta Coraglia
University of Milan

News

Formal Modeling of Ignorance, Urbino 7-8 September

The two-day workshop *Formal Modeling of Ignorance* took place on the 7th-8th of September in Urbino, as a satellite event of the SILFS Triennial Conference, and as a meeting related to the project *Developing Kleene Logics and their Applications* (DeKLA). It was organized by Stefano Bonzio (University of Cagliari) and Pierluigi Graziani (University of Urbino).

Although the notion of "ignorance" was discussed already by Socrates and is as old as that of knowledge, the former has attracted less attention from philosophers than the latter. This tendency has been partially changing in the last years, thanks to the publication of an increasing number of logical and epistemological studies (two representative examples are van der Hoek & Lomuscio 2004 and Peels 2023). The workshop brought together experts of the formal (logical) modeling of ignorance. It featured five invited talks.

Alessandro Aldini (University of Urbino)'s opening talk "On the modeling and verification of the spread of fake news, algebraically" (for the paper, see here) presented a model for the formal analysis of the spreading of fake-news within a network of social agents. The model, substantially based on process algebras, allows to describe the dynamics of spreading as well as the rate speed of the propagation. Three-layers are designed for modeling social agents in isolation (level 1) and in presence of other agents (level 2), and the entire social network where agents can behave as "believers" or "fact-checkers" (level 3). Although the notion of ignorance is not explicitly modeled, it is a crucial element, as some agents in the formal model are ignorant about a fake-news while others are not.

Mirko Tagliaferri (University of Urbino)'s "Studying the map: a taxonomy of formal models of ignorance" proposed an up to date summary of the varieties of different notions of ignorance that have been logically investigated. A first hand distinction regards authors approaching ignorance as a derived concept with respect to knowledge (e.g. Fine 2018, Fano & Graziani 2021) and those renouncing to the connection with knowledge and preferring a primitive formalization (e.g., van der Hoek & Lomuscio 2004, Bonzio et al 2023). The talk explored in details the treatment of "ignoring that", logically rendered as $\neg\mathbf{K}\varphi$, of "ignoring whether" ($\neg\mathbf{K}\varphi \wedge \neg\mathbf{K}\neg\varphi$), and the notion of Steinsvold's (2008) "A Note on Logics of Ignorance and Borders" in the setting of epistemic logic. The author finally argued in favor of the investigation of a new form of ignorance, employing a modal logical setting featuring operators for both beliefs and justified evidence. In such a context, an agent's ignorance could be analysed using a fine-grained distinction of the reasons why ignorance is present (cf. Tagliaferri 2023).

The remaining three-talks form part of the same research enterprise, and were presented in the form of a trilogy.

Mattia Petrolo (University of Lisbon) gave a talk titled "A New Modality". In the first part, he introduced the two main epistemological perspectives on ignorance, namely the Standard View (SV) and the New View (NV) (LeMorvan& Peels 2017). In the second part, he discussed the ongoing debates in the field regarding formal models of ignorance, presenting three operators and their respective systems for ignorance-whether (not knowing whether) (van der Hoek & Lomuscio 2004), ignorance-that (unknown truth) (Steinsvold 2008), and disbelieving ignorance (Kubyshkina & Petrolo 2021, Gilbert et al 2022). Comparing these three operators with the epistemological analyses described in the first part of the talk, Petrolo showed the existence of a gap between epistemological analyses and formal models of ignorance. Petrolo also highlighted that the formal representation of disbelieving ignorance is not interdefinable with usual K or B operators on standard frames. Petrolo concluded that this evidence might point to the epistemological hypothesis that ignorance (at least in some of its forms) is not reducible to other mental attitudes, such as knowledge and belief, and this might point to the development of an ignorance-first approach.

Marianna Girlando (ILLC Amsterdam)'s talk was titled "Knowledge strikes back". In the first part, she emphasized how, precisely for the reasons highlighted by Petrolo, it is very interesting to formalize disbelieving ignorance using proof-theoretic tools. Specifically, she pointed out how it is intriguing to define classes of models and formal systems that model both ignorance and knowledge and satisfy the following three desiderata: to capture at least all three notions of ignorance; to model an S4- or S5-notion of knowledge; and the formal systems should be analytic (sequent calculus). Following this research direction, in the second part of her talk, Girlando proposed both a Hilbert-style calculus ($\mathcal{H}_{WUDI}$) and an equivalent labeled sequent calculus (*lab*WUDI) that include the three operators described by Petrolo. She demonstrated metatheoretical results in the context of this second system, such as soundness, completeness, and towards completeness the termination of proof search. Girlando concluded her presentation by suggesting some future research directions, including an unlabeled sequent calculus for disbelieving ignorance.

The third episode of the saga was presented by Ekaterina Kubyshkina (University of Milan)'s talk "Return of Ignorance". In the first part, after analyzing the four types of ignorance presented by Peels (2014), namely Disbelieving Ignorance, Suspending Ignorance, Deep Ignorance, and Warrantless Ignorance, Kubyshkina argued for the thesis that only Disbelieving and Deep Ignorance constitute fully Excusable Ignorance. In the second part, Kubyshkina provided a sound and complete logic for this kind of ignorance, called the Logic for Excusable Ignorance (LEI). The logic is characterized by a relational semantics with truth-value gaps that allows the distinction of cases of Disbelieving and Deep Ignorance from other cases of ignorance. In the last part, Kubyshkina extended LEI to a Public Announcement Logic (PAL) by introducing an original update operator. The novelty of this operator consists in the fact that instead of eliminating epistemic possibilities as usual in PAL systems, it creates new ones. Finally, Kubyshkina proved that this extended logic is sound and complete. This new system permits one to capture the change of an agent's epistemic state in a dynamic setting and describes the conditions under which excusable ignorance can be transformed into non-excusable one.

At the end of the workshop, the organizers suggested establishing a network that would unite all of the scholars in the field of formal modeling of ignorance, so as to encourage and facilitate regular discussions about new and challenging problems and their possible solutions. This proposal received unanimous support from all of the speakers.

PIERLUIGI GRAZIANI
University of Urbino
STEFANO BONZIO
University of Cagliari

THE GOVERNMENT FINALLY DECIDES TO PUT AN END TO ALL THE ARGUMENTS.