The Reasoner

Volume 12, Number 4
April 2018

thereasoner.org
ISSN 1757-0522

Contents

Editorial

Dear Reasoners, it is my pleasure to introduce you to Richard Zach. Richard is a logician with a broad range of interests, in addition to being one of the main forces behind the well-known Open Logic Project, openlogicproject.org. I would like to thank Richard for his time, and on behalf of the reasoning community, for the tremendous job he's doing for the wider understanding of logic. You can learn more about his work and interests on his very rich webpage richardzach.org/

Hykel Hosni
Università di Milano

Features

Interview with Richard Zach

Hykel Hosni: Can you please tell us something about your background?

Richard Zach: I started undergraduate work at the University of Technology, Vienna, in computer science and at the University of Vienna in formal logic. Two years into my studies, I combined the mathematical and theoretical computer science aspects of my interests in a self-designed major, "Computational Logic." This required an application to the Federal Ministry of Education, but was eventually approved. One of my co-conspirators was Helmut Veith, who went on to a distinguished but tragically short career in formal verification. I did most of my work at the University of Technology, working in proof theory with Matthias Baaz and Chris Fermüller and automated theorem proving with Alexander Leitsch. I wrote on proof theory of finite-valued logics for my master's degree, and continue to collaborate with the logicians in Vienna on work in proof theory and non-classical logics.

HH: Quite a line-up of great logicians you had in Vienna! With such a start, no wonder you wanted to continue for a PhD.

RZ: Yes indeed! For my PhD, I moved to the States, to the Group in Logic and the Methodology of Science at Berkeley. That program was founded by Tarski, and aimed to bring together mathematicians and philosophers interested in logic, foundations, formal methods. It continues to be a thriving and inspiring program. Unfortunately, right after I arrived, budget cuts forced a wave of retirements, and many of the faculty in the math department I had hoped to study with were no longer there. I still had a wonderful time and learned a lot. I took a course on proof theory, nominally with Jack Silver, although it was actually taught by Jeremy Avigad who was just finishing up his degree in the math department. The two-course graduate metamathematics course was taught by Hugh Woodin in the first term, and Joel Hamkins in the second. I needed to take courses in philosophy as well, and started with theory of

knowledge taught by Barry Stroud and modal reasoning with Charles Chihara. I would go on to learn a lot of philosophy of math from Chihara, but Stroud's course really got me interested in doing work in philosophy. I studied Kant with Hannah Ginsbourgh, Frege with Hans Sluga, philosophy of language with Ernie Lepore, Donald Davidson, and Stephen Neale.

HH: That is serious mathematics *and* serious philosophy ...

RZ: A year later Paolo Mancosu was hired by Berkeley, and I started to work with him on history of logic and the Hilbert program very soon after that. I also continued to pursue my interests in computer science, studying complexity theory with Christos Papadimitriou and Umesh Vazirani. I would also meet with Jack Silver regularly, though: he was the only one interested in proof theory, although he didn't work on it himself. He and Mancosu became my supervisors, although Paolo did most of the supervising. After a brief stint at Stanford, I was hired at the University of Calgary in the philosophy department, mainly to teach logic to computer science students.

HH: Can you recall the spark that ignited all this?

RZ: Yes! I picked up *Gödel, Escher, Bach* as a teenager.

HH: Well, that book made many of us logicians!

RZ: Indeed. I was fascinated by the work of Gödel and Turing. As a student, it was really logic all the way – first in the foundations of mathematics and automated theorem proving, then pure proof theory and non-classical logic. At Berkeley I added interests in the history of logic and philosophy of mathematics. I continue to work in all these areas, and have branched out to the history of analytic philosophy more broadly. But I stick to the logicians, especially Carnap recently.

HH: What about Carnap, in particular?

RZ: Well, my major project recently has been the Collected Works of Rudolf Carnap – fifteen volumes of all of Carnap's writings, with revised or new translations of everything not written in English. That's a major undertaking, as you can imagine, and it's taken more than a dozen people and almost two decades, but the first volume should finally appear this year. I'm involved with three later volumes.

HH: Can you give a quick preview to the readers of The Reasoner of what this editorial effort will bring about?

RZ: Carnap's work of course is of immense importance to logic, the philosophy of science, semantics, and probabilistic and inductive inference. Although many of Carnap's later writings have been available in English, this is not true of most of what he wrote before the 1930s. That includes his early work on the philosophy of science and mathematics in the 1920s, with the exception of the *Aufbau*, and almost all of his contributions to logic in the 1920s – those will be translated for the first time in volumes 1 and 3. Much of his work from the 1930s, including his central contributions to mature logical empiricism, is scattered and translations are not uniform. They will be collected in volumes 4 and 6, which I'm responsible for. The project has a website rudolfcarnap.org.

HH: Well, this will certainly be of interest to many reasoners across disciplines. But let me go back to your early academic history. Up to the PhD, your path doesn't seem to fit comfortably in any of the current institutionalised "programmes". Do you think this is due to chance, or is it rather a feature of the truly and intrinsic interdisciplinary nature of logic?

RZ: Both the logic program at the University of Vienna and the Program in Logic and the Methodology of Science at Berkeley were/are institutionalized. But they are special cases. There are very few other programs that build bridges between disciplines around logic. So, as far as I'm concerned, it was certainly a measure of chance that allowed me to pursue the path I eventually chose. It is certainly also true that the interdisciplinary nature of logic is not conducive to a proliferation of specialized logic programs. Or rather, it is not conducive to *un*specialized logic programs, which train people in a number of disciplines to apply logic there. You'd need to offer training not only in logic but also all the other areas where logic is applied. So it really only works well in graduate programs where the students already have a foundation in at least one other discipline.

HH: That certainly makes sense. So how do you think your training reflects in your current research?

RZ: It has enabled me to do some things that I probably wouldn't have been able to do otherwise. It's hard to do philosophy of mathematics well if you don't have mathematical training and experience. My own training was in proof theory, so that's served me well in working on, say, Hilbert's program, where formalization and proof plays a central role, or on Carnap's work in logic. It also allows me to talk to different audiences, both in research but also in teaching.

HH: Please, tell us a bit more about that.

RZ: Most of the students in my introductory logic courses are computer science majors. It helps to be able to talk about the uses of logic in knowledge representation, specification, verification, and automated deduction. I'm trying to bring this perspective to the *Open Logic Project*: logic is a mathematical discipline, but it has applications to and relevance for all these other areas, and your typical text doesn't stress that enough – or stresses it only for one area. In the research literature as well there's a fair bit of reinventing the wheel; there's value, say, in philosophers looking at what the computer scientists are doing with modal logic.

HH: I certainly agree. The *Open Logic Project* is internationally well-known and widely used. Can you tell us how the project started and where it's heading?

RZ: I'm teaching a few intermediate logic courses regularly. Our second logic course covers the metatheory of first-order logic, including completeness and undecidability. The students who take it aren't mathematics students; most of them are computer science or philosophy students. Even the computer science students don't have a lot of mathematical training: they don't yet know how to read or write mathematical proofs. Most advanced logic textbooks assume that, however. That makes it hard to use one of the usual texts. In the past, I have used Boolos, Burgess, and Jeffrey's Computability and Logic. It's not a very mathematical text, but still too quick.

HH: So you decided you would do your own text ?

RZ: Yes. My colleague Nicole Wyatt and I decided of writing one approaching the material more gently and including a lot of material that isn't typically covered, such as basic set theory and proofs by induction, and not assume any mathematical background. Rather than do this with a publisher, we decided

to give it away for free. That created the opportunity to start a larger project: a modular collection of textbook materials that others could use as well, and perhaps expand.

HH: And that led to the GitHub platform github.com/OpenLogicProject/OpenLogic

RZ: Eventually, yes. I recruited a few other colleagues to advise and contribute. Aldo Antonelli gifted his lecture notes on model theory and modal logic, and Jeremy Avigad his on computability theory and incompleteness. I cut those notes into modular chunks, made all the notation and terminology uniform, and added a lot of material, both to the existing notes and sections written from scratch. Nicole and I received some funding from the University of Calgary and the Government of Alberta, which allowed us to hire some very smart students. Dana Hägg and Samara Burns helped edit the text, expanded additional notes from Nicole an Audrey Yap, and wrote a fair amount of new material that wasn't covered in Aldo and Jeremy's notes. The result is now an online repository of expository logic material.

HH: It's not just a big PDF that people can download for free, then.

RZ: Not really, it's all the LaTeX source code. Since it's modular, you can mix and match the sections to build individualized textbooks. I've so far made two: Sets, Logic, and Computation is the text for the second logic course. It introduces students to the basic notions and results of naive set theory, and in the process trains them in how to read, organize, construct, and write proofs. In the second part, it talks about first-order logic: syntax and semantics, theories and models, proofs (using sequent calculus and natural deduction as alternative systems), completeness, compactness, and Löwenheim-Skolem theorems, and finally Turing machines and undecidability. Nicole and I have now used it for that class for a few years, and our students like it a lot better than the previous text. (We've in fact run surveys both in courses that use the new text and in some that use the old for comparison.) I have also made a textbook on computability and incompleteness from it for the third logic course in our sequence, and am now working on a textbook on modal logic.

HH: Do you think the Open Logic Project could serve as the basis for a wider project aimed at the general (non academic) public? Wouldn't it be fantastic if the international and multifaceted logical community contributed to creating a sort of "theoretical minimum for good reasoning" that citizens, journalists, and policy-makers could freely access to?

RZ: That would be fantastic! It has to be said though that the Open Logic Project is still a mathematical text, even if it is aimed at an audience who doesn't necessarily have a mathematical background or mathematical interests and motivations. It would be hard to make it serve the other function you mention: journalists and policy-makers don't have time to read through 100 pages of naive set theory in order to understand the model-theoretic definition of consequence, say. But that just means that you shouldn't wait on the Open Logic Project to develop open learning materials on good reasoning generally. We already have good open textbooks on what's typically taught in a first-year introduction to formal logic, where the emphasis is on symbolization, semantics, and formal proofs. P. D. Magnus's *forall x* has been around for over a decade, and because it's open, people can use it as a starting point for derivative versions with different features. There are now at least eight of them; one combines it with the open "Introduction to Reasoning" by Cathal Woods, and a version by Edward Elliott includes a part on probability theory. We should also develop open learning materials on cognitive bias, statistical reasoning, philosophy of science, social epistemology, etc. etc.

HH: You've done an amazing amount, and yet much is still to be done! Can you tell us about your current research projects?

RZ: I have a few projects I'll have completed soon. The first is the Carnap project we talked about earlier. Then, with Paolo Mancosu and Sergio Galvan I'm working on an introduction to proof theory. We are taking a historical approach, and hope to make proof theory accessible to philosophers. There's a paper on the development of the semantics of first-order logic in the 1920s and 1930s that's waiting to be sent off. I hope to pull together the results of my historical studies over the last decade in that. I'm also excited about the philosophy of code. During my last sabbatical I sat in on a course by Brigitte Pientka at McGill, and that rekindled my interest in theory of programming languages. There are obvious connections to my interests in history and proof theory. Although philosophers have long concerned themselves with computability generally, they haven't really engaged with what computer scientists do all that much, at least not to the extent that, say, philosophers of science care about scientific practice and the history of science. That's changing, and I think it's an exciting development, and logicians will play an important role in it.

News

Kent Formal Epistemology Conference, 18-19 December

The blustery hilltop campus of the University of Kent at Canterbury was the venue for a two-day formal epistemology conference hosted by the multidisciplinary Centre for Reasoning. Eight speakers presented in an informal workshop-like atmosphere, which proved fertile ground for discussion among the experts while allowing non-specialists an insight into the state of the art.

After the first coffees of the day, Catrin Campbell-Moore (University of Bristol) discussed some self-referential probabilities that produce unstable beliefs for a reflecting agent. Drawing on Kripke's work on the semantic paradoxes, the speaker argued that these self-undermining beliefs are best accommodated in a framework of supervaluationist logic, with belief states modelled as sets of probability functions rather than as precise probabilities. This would seem to recommend suspension of judgment in the case of evidential dependence and - more controversially - metaphysical vagueness in the case of causal dependence.

Some small debt to Kripke was acknowledged also by Bernhard Salow (University of Cambridge) in his talk about puzzles of dogmatism and belief revision. Using some motivating examples, Salow argued that we should reject the principle of rational monotonicity (RM) which states that an agent who learns some fact consistent with her existing set of beliefs should continue to hold those beliefs. Doing so creates fruitful logical space for a principle of conditional dominance (CD); if you believe that A is better than B *on the condition that* A *and* B *are not equally good*, then go with A rather than B. Salow argued convincingly that once RM is rejected, CD can explain why we are not obligated to avoid sources of evidence that contradict

beliefs that we hold to be true.

Julia Staffel's talk entitled 'Non-ideal rationality and the problem of second best' raised the question of how we should measure divergence from the ideal credences if rationality requires more than probabilistic coherence alone. Staffel (Washington University) distinguished two kinds of approach. A bundle strategy measures the distance between one's credences and the closest function fulfilling all requirements simultaneously; a piecemeal strategy measures how closely the credence function approximates fulfilment of each requirement individually and then aggregates. It was demonstrated that for either choice of strategy, if one rationality constraint is necessarily violated then there is no guarantee that the optimal available state fulfils all of the remaining requirements - an instance of the phenomenon known to economists as the *theory of the second best.*

'Epistemic Optimism' was presented by Julien Dutant (King's College London) as a response to the New Evil Demon (NED), continuing the research project of knowledge-first epistemology promulgated by Williamson. To be *epistemically optimistic* is to hold that it is rational to believe P iff you are not in a position to know, that you are not in a position to know P. The speaker argued for a novel position of *Global Epistemic Optimism* based on a two-tiered theory of knowledge and rational belief, which allows for defeasibility and internalist-friendly judgments when faced with NED. Some work-in-progress on a formal logical framework was also presented. A final lively Q&A session brought the day to a close, though naturally the discussion carried on well into the evening.

The second day began with Ginger Schultheis' talk on 'Belief and Probability'. The epistemic Bayesian was faced with a counterexample for simple Lockean Supervenience accounts of conditional probabilities and belief revision. A probabilistic analogue of RM (argued against in Salow's talk) was again diagnosed as a source of problems. The speaker presented the case for imprecise Lockeanism, with a set of probability functions representing a subject's belief state and each member being updated by conditionalization. A conservativeness constraint can be imposed on this credal committee; let t be some threshold such that S believes P just when each member of the set has $Pr(P) > t$. This was seen to accommodate the initial counterexample and to support the general approach of reducing belief to subjective probabilities.

Cat Saint-Croix (University of Michigan) gave a thought-provoking presentation on 'Immodest Modesty and the Epistemic Good'. An agent's views on epistemic values like certainty, accuracy, conciliation and belief-revision may be reflected in a choice of epistemic utility function or scoring rule. But how might one model the case where an agent is uncertain about what it is to be epistemically good? The speaker proposed that, faced with such normative uncertainty, one should take the expectation over expected value returned by the various candidate utility functions. This allows for creditworthy epistemic agency and a kind of modesty with respect to judgment about what it is to be rational, while vindicating the widely recognised rationality constraint of *epistemic immodesty* - that is, a scoring rule should guarantee that its adherents view their own (rational) credences as the best possible.

Drawing on work from McGee (1985), Reuben Stern (MCMP) presented some instances of nested indicative conditionals in natural language which appear to be classically (truth-functionally) invalid. But when viewed from a Bayesian perspective, it was argued, such examples actually illustrate an important normative feature of inference; specifically, that if one becomes convinced of the premises of a modus ponens-type argument then one should also become convinced of its conclusion. On this reading, such cases should be taken not as an argument for the invalidity of modus ponens as a general rule, but rather as a lesson about synchronic vs diachronic aspects of inference in natural language.

The final paper of the conference was given by Ben Levinstein (University of Illinois at Urbana-Champaign). 'Cheating Death in Damascus' provided summaries and entertaining counterexamples for the leading contenders in the theory of rational action (incorporating what is presumably the first utterance of the phrase "schlepping over to Aleppo" in the context of epistemic decision theory). After demonstrating the failings of Evidential and Causal Decision Theory, the speaker outlined the novel Functional Decision Theory (FDT) that he is developing in collaboration with Nate Soares (Machine Intelligence Research Institute). FDT was seen to outperform EDT and CDT in situations where an accurate predictor (the Death of the title) has access to one's decision-making procedure. It was noted by the speaker that in its reliance on subjunctive dependence, a fully-developed FDT will require a theory of non-trivial counterfactuals allowing for logically impossible antecedents.

This conference was organised by Dr Jason Konek, formerly of University of Kent. Kent's Philosophy Department and the Centre for Reasoning would like to take this opportunity to wish Jason the very best in his new position at the University of Bristol.

Gavin R. Thomson
Philosophy, Kent

Calls for Papers

E. W. Beth Dissertation Prize, 2018: awarded by the Association for Logic, Language, and Information to the best dissertation which resulted in a Ph.D. degree awarded in 2017, deadline 23 April.

Non-Classical Modal and Predicate Logics: special issue of *Logic Journal of the IGPL*, deadline 30 April.

Pluralistic Perspectives on Logic: special issue of *Synthese*, deadline 1 June.

What's Hot in ...

(Formal) Argumentation Theory

The study of computational argumentation has made significant advances since the work of P.M. Dung and others in the mid nineties. You may recall from the overview in the first (June 17) edition of this column, that a Dung argumentation framework (AF) is essentially a directed graph in which the nodes denote arguments, and a binary relation between nodes denotes that one argument is a counter-argument to ('attacks') another. Two key areas of study have then focussed on: i) applications of Dung's theory of argumentation to the formalisation of non-monotonic reasoning and de-

cision making by individual agents and multiple agents engaging in dialogue; ii) identifying subsets – or so called 'extensions' – of the given arguments in an AF that are said to be justified, or 'winning', under various semantics that differ in their criteria for membership of arguments in extensions.

An important development towards practical applications of argumentation has been the inauguration of the First International Competition on Computational Models of Argumentation (ICCMA) in 2015, and the subsequent 2017 edition of ICCMA (see AI magazine 37 (1), 102 and argumentationcompetition.org/2017/ respectively). Deciding which sets of arguments are extensions of an AF and whether a given argument is in some or all extensions, under a chosen semantics, is computationally difficult. Applications utilising argumentation therefore require efficient 'argument solvers'. To promote the development of such solvers, ICCMA invites researchers to submit the results – the times taken to answer the aforementioned decision problems – for their solvers, assessed on a series of challenging benchmark AFs.

The ICCMA competitions have provided valuable impetus for the development of efficient argument solvers. However, I would advocate complementary assessment of solvers that address the needs of applications in which arguments are structured chains of reasoning from premises to claims in some formal logical or informal natural language. The ICCMA competitions assume AFs consisting of arguments that have already been constructed and related according to whether they attack each other (e.g., because the claim of one argument negates the premise, or conclusion of a rule, in another argument). However, argumentation as applied to reasoning and decision making often involves submitting a query to a knowledge base, then constructing arguments in support of (i.e., that conclude) the query, and then constructing arguments that attack these supporting arguments, and then arguments that attack these attacking arguments, and so on. We thus witness argument game proof theories and dialogical formalisms that begin by issuing an argument X, and then successively deploying arguments that attack their predecessors, so generating a tree of arguments rooted in X. These use contexts invoke a series of computationally challenging tasks, such as: 1) find all proofs (*qua* arguments) from a knowledge base that conclude a given claim; 2) For a given argument X, invoke the task in 1) so as to find all arguments whose claims negate a conclusion or premise in X (i.e., all arguments that attack X). Hence, assessing these tasks against benchmark knowledge bases expressed in different formal and informal languages might fruitfully stimulate appropriate implementational efforts. Moreover, the ICCMA benchmark AFs for which the decision problems are computationally highly demanding, are those that contain cycles. However cycles may not be an issue for argument game proof theories and dialogues, since these typically prohibit repetition of arguments so that the successive identification of attacking arguments yields *finite* trees of arguments that can be efficiently processed to determine whether the root argument is justified under a given semantics.

Sanjay Modgil
Informatics, King's College London

Medieval Reasoning

I realised that in the last few months this column has almost exclusively mentioned authors belonging to the Latin tradition. This month I am going to spend a few words on Ibn Sina, i.e. Avicenna (980-1037), lest I present an incomplete and therefore misleading picture of medieval reasoning as an exclusively later medieval, Latin speaking and European matter.

While Avicenna's enormous influence on both Arabic and Latin philosophy has been well known for a long time – for example, his *al-Qanun fi al-Tibb, The Canon of Medicine* was used as a textbook in European universities up until the 18th century – only in relatively recent years have Western scholars started working systematically on his philosophy. I would like to introduce an interesting feature of Ibn Sina's account of intuition, that is (as far as I know) unsual yet relevant for medieval theories of knowledge and of ordinary reasoning. A long tradition from antiquity to post-Kantian aesthetics, epistemology and philosophy of science – despite significant variations – seems to have equated intuitive and non-discursive knowledge, roughly as follows: overall, intuition has been treated as a privileged, peculiar or extraordinary epistemic state, associated with artistic, creative and mystical experiences in which some complex understanding or construction spontaneously pops into being. Insofar as some kind of non-discursive immediate and intuitive grasp of some articulated fact or deeper truth (whatever that might be) seems to be involved, the Kantian "intuitive understanding" (anschauende Urteilskraft) or "intellectual intuition" – at least as they have been received by the post-Kantian tradition – have never appeared to be too far from the "intellectual vision" (visio intellectualis) or the "spiritual intellect" (intellectus spiritualis) that you can find in authors like Gregory the Great or in the 13th and 14th century discussions about prophetic knowledge. However, the "anschauende Urteilskraft" might turn out to be further away from these latter discussions than from earlier accounts like Gregory's, exactly because of Avicenna's impact – which has yet to be fully assessed. In the *Kitab al-Najat, Book of Salvation* or *of Deliverance* (known to the Latin world as *Liberatio*), Avicenna makes some specific remarks about intuitions:

> If a person can acquire knowledge from within himself, this strong capacity is called 'intuition'. [...T]he intelligible truths are acquired only when the middle term of a syllogism is obtained. This may be done in two ways: sometimes through intuition, which is an act of the mind by which the mind itself immediately perceives the middle term. This power of intuition is quickness of apprehension. But sometimes the middle term is acquired through instruction [...]. It is possible that a man may find the truth within himself, and that the syllogism may be effected in his mind without any teacher.
> *Kitab al-Najat* II.6

While for Avicenna 'intuition' is some ability that some pos-

sess in a higher degree than others, it seems to be a feature of the ordinary workings of the human intellect; in this picture, intuition is not meant to give access to any exclusive revealed truth that cannot be acquired by any ordinary means – namely instruction. In other words, intuition might be a special attitude, but not *that* special. Although intuition is an immediate grasp of something – and therefore *per se* is not discursive – Avicenna claims that it is not the grasp of a conclusion or of some complex non-propositional or non-argumentative understanding, but of the correct middle term to insert in a syllogism. What we have intuition of is a term shared by two categorical sentences arranged in such a way that, by discharging that very term, we can draw a conclusion s.t. its predicate is the remaining term of the major premise and its subject is the remaining term of the minor premise. Leaving aside the syllogistic jargon, the claim is that intuition is about connections between statements (and possibly about the things in the world verifying those statements); the intuitive grasp of these connections allows us to reach a conclusion, by following the usual inferential patterns. In other words, intuition is neither of some complete (supposedly true) statement nor of a conclusion that we wouldn't be able to track back to its premises: instead, for Avicenna 'intuition' is of a term and it does not take us out of the ordinary argumentative processes, rather it contributes to them. Overall, on the one hand, a reconsideration of those later Latin theories of intuitive knowledge more or less explicitly referring to Avicenna could be in order: it might turn out that they are of a more discursive and argumentative persuasion than their 18th century distant cousins. On the other hand, Avicenna's account seems to be an intuitively (!) sensible take on how intuition works, at least for those of us who are not mystics or creative geniuses – if there are any such things.

Graziana Ciola
UCLA

Uncertain Reasoning

The robots are going to take over! It's been in the news recently that various serious people have suggested we need to take the threat posed by possible future superintelligent machines more seriously than we do. Stephen Hawking, who died this week, was among those worried by this threat, as is Elon Musk. Not all tech billionaires are equally scared of Skynet: Mark Zuckerberg has been particularly bullish about the prospects for AI making our lives better. Nick Bostrom recently wrote an extensive and careful treatment of the risks of superintelligent machines (Superintelligence, OUP, 2014). The book, however, stays on the level of "this could happen" and "this would be costly", rather than trying to quantify the costs and probabilities associated with such events. One might think that I am bringing this up in order to critiicise that lack of precision, but in fact, I think this is a case where that lack of precision might be the best approach.

When dealing with uncertainty, it often seems like the right approach is to try to quantify that uncertainty. We have a huge literature – that spans many academic disciplines – on how to do that. Likewise when assessing the costs or benefits of some possible event or action. However, it seems to me that there are limits to what we can (and should) quantify. In medical ethics, for example, there is a debate about the morality of putting an economic value on a human life.

What I want to suggest here is that perhaps "low probability high cost" events (like a malevolent superintelligence) are another case where we should resist the urge to quantify. Let's take the risks posed by a possible future unfriendly superintelligence as an example. How costly might such an occurence be? Well, in the worst case, the powerful superintelligence destroys everything in the universe to process it into more computing machinery for itself. How do we quantify the costs of such an event? Leave aside what I said above about the difficulty in quantifying the disutility of the loss of a human life, and we still have the problem that the *extinction* of the human race seems significantly worse than just most people dying. Put that aside and think that such a machine would also destroy any other possible life in the universe. That also seems bad, but also hard to put a number on. So, how bad is this possible future? Is it a billion euros of damage bad? Is it 10^{18} euros bad? And how likely is this unfriendly AI scenario? Again, it seems impossible to put a number on it. Let's imagine that it's pretty low, but not zero. Is it a one in a million occurence? One in a billion? So how much should we spend to avoid such a scenario becoming reality? The standard way to answer this question is to (somehow) produce probability and utility estimates, and work out the expected cost, and then say that we should pay up to the expected cost to prevent the event. But in a case where the probability of the outcome is very low, and the cost is very high, how much we should spend on prevention varies wildly as estimates of the probability and utility vary within reasonable bounds. If we take the optimistic values for the probability I threw out earlier, we should spend one euro to stop it. If we take the pessimistic ends of the guesses, we should spend a trillion euros to stop it. Who's to say which of those guesses is better? Has it actually helped to quantify this decision problem? We are no closer to a decision than we were! So in these sorts of circumstances, it's simply not worth engaging in quantification. Other, less quantitative methods of decision making should play a role in this sort of case.

Let's turn to a different aspect of this issue. It's another case mentioned by Nick Bostrom (Pascal's Mugging, 2009, Analysis Vol 69 No 3). The basic idea is that take some possibility that is extremely costly (for example, an asteroid hitting the earth and causing untold damage). Now, I presume you think it extremely unlikely that I can prevent such a disaster on my own, but let's imagine that you think there is some tiny small chance that I have some sort of power that would allow me to prevent it. Then, there is some small positive amount of money that you should be willing to give me if I promise to do what I can to prevent the bad outcome. But if I come up to you in the street and say "I promise to prevent asteroid attacks if you give me a euro", I doubt you'd give me the cash. This example relies on your having a small but non-zero probability for my having this power. But why should you have that? It seems you have good reasons for placing an upper bound on how likely you think such a possibility is, but there's no good reason why you ought to assent to any particular lower bound on that probability. If you refrain from fully quantifying your probability here, then the argument doesn't go through. (Decision making

with "imprecise" probabilities is a big topic and what does actually follow is up for debate, but at the very least, one version of the "Pascal's mugging" problem is solved).

A lot of work on uncertain reasoning concerns how to quantify uncertainty; I think we should perhaps think a little more careful about when to quantify uncertainty.

Seamus Bradley
Philosophy, University of Tilburg

Mathematical Philosophy

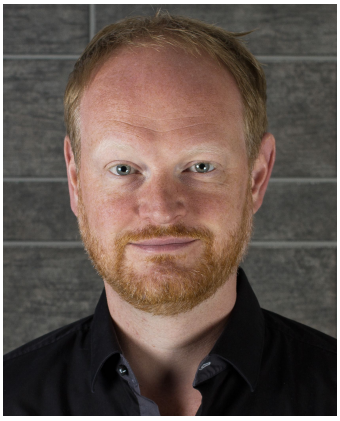
I'm going to use my space this issue to talk about formal or computational models. In the Neurophilosophy program at the Graduate School of Systemic Neurosciences (GSN), where I'm a PhD candidate, we're interested in building bridges between neuroscience and philosophy by using the tools and results of one discipline to examine questions in the other, in both directions. I came through the Master's program in Logic and Philosophy of Science at the MCMP, with a strong focus on Bayesian methods, and did some work on Bayesian systems of argumentation. I'm now looking to take this formal apparatus and apply it to contemporary work in the cognitive sciences.

An accusation that is sometimes raised against philosophers who use Bayesian or probabilistic formalism in their work is that they're constrained or misled by their over-reliance on a methodological Maslow's hammer. To a Bayesian, the argument goes, every problem looks like it would be amenable to a probabilistic solution.

To respond to this argument, we often point to examples from the sciences that show the success of applying Bayesian formalisms to different areas of investigation, in perhaps unexpected ways. In fact, I think it's possible to do more than just point to success stories – I'll say more on this a little later.

An excellent example of how computational modelling is used in psychiatry can be found in Huws, Vogelstein and Dayan (2009), where the authors use a normative Bayesian decision-theoretic framework applied to a decision-making activity to predict important differences in reasoning caused by depressive disorders. For some critics, this application of decision-theoretic apparatus in psychiatry may seem surprising. The results of their modelling, however, are compelling. The authors are interested in how two central concepts in the diagnosis of depression, helplessness and anhedonia, can explain typical behaviors associated with depression patients. Their choice of computational model allows them to formalize these two concepts. Specifically, helplessness is formalized in their modelling language as a simple prior over the outcome entropy of actions in uncertain environments; that is, learning that action A leads to outcome B does not raise the probability that action A will lead to outcome B in the future. Anhedonia is formalized as a problem with reward sensitivity: a failure to correctly classify the utilities of outcomes. The authors apply this formalism to a reinforcement learning task with uncertain outcomes. The same reinforcement learning task is also presented to test subjects: an experimental group of patients that have been diagnosed with Major Depressive Disorder, and a control group of mentally healthy individuals. The results in the paper show that the output of the formal model of the decision-making process accords closely with the responses from the experimental and the control groups. Based on this, the authors suggest that the two symptoms, helplessness and anhedonia, could explain much of the behavioral output of patients with depression, and therefore should be the target of clinical treatment.

For a philosopher of science, this paper gives a lovely example of the methodology of modelling in the sciences: First, a modelling language is chosen and applied to the target domain. In this case, a Bayesian decision theoretic framework that deals with utilities and probabilities is applied to a reasoner's representation of outcomes in uncertain situations. The model can then compute 1/ How these representations are updated over time as new information is learned, and 2/ How these representations are used to guide decisions. Next, a characteristic of interest in the target domain is given a formal counterpart in the modelling language; i.e. the two symptoms outlined above. The model is then used to predict behavior on a set scenario, in this case the differences in decision-making between experimental and control groups, and this output can be compared with the results of experiments in the target domain. The authors use the success of the model's predictions to formulate a hypothesis as to the explanatory power of the two factors they were examining in their experiment.



The analogy with modelling methodology in the physical sciences is clear: This is very much the same methodology that is followed when using calculation to predict the movement of charged particles in a magnetic field in a physics experiment, or the results of a chemical reaction in chemistry class. However, there is a significant area where this analogy does not hold up, and that is in the nature and ontology of the target domain. In the physical sciences, there is wide agreement on what the objects in the target domain are, as well as their characteristics and ontological status. The cognitive sciences, on the other hand, offer a multitude of levels of description of the target system: the mind. While each individual experiment is often very clear about the level of description that is relevant to the question under discussion, larger questions remain: Are these different modelling activities addressing the same target system? To what extent are the objects in the target domains that are represented in different modelling experiments the same objects? What are the broader consequences of the ontological assumptions made in such experiments, and are these assumptions broadly coherent? The goal of my research project is to render explicit the sometimes-implicit assumptions made in experiments like these in the cognitive sciences regarding the objects of the target domain, in order to compile a more complete and coherent view of what the inhabitants of this domain are like.

These are higher-level questions of scientific methodology that can shed some light on the justification for applying a computational framework to questions in the cognitive sciences; justification that can be stronger and more formally compelling than simply pointing to success stories from contemporary research. Using lessons from philosophical work examining and codifying model use in the hard sciences, it will be possible to say more about the applicability of a probabilistic framework in the cognitive sciences; for example, in modelling how depression

affects reasoning. After all, Maslow's hammer argument isn't a worry if you're using a particularly flexible hammer.

Harry Waterstone
Munich Centre for Mathematical Philosophy

Philosophy and Economics

This is how it goes sometimes. One moment you write a column in the Reasoner that is highly sceptical of classic event formats such as the larger conference or the smaller thematic workshop (in February 2018, to be precise). The next thing you know, you are participating in a wonderful workshop that is – at least that is how it looks like from the outside – organized in such a classic format: meet up in the early afternoon, put up five speakers from junior to senior for slots of 45mins each, discuss, drinks, dinner. That is what happened to me in March.

A congenial band of researchers at the Tilburg Centre for Logic, Ethics, and Philosophy of Science (TiLPS) – Huub Brouwer, Bart Engelen, and Naftali Weinberger – put together a workshop on the theme "Making Hard Choices: The Ethics and Economics of Health Care". They invited philosopher-economists to discuss their work and explore the joint theme. And indeed, what is rare in workshops – even when they are as nicely defined as this one – is that there were indeed several strands of very close connections between the talks: each speaker conceptualised the ethical and economic implications of 'hard choices' concerning health care slightly differently.

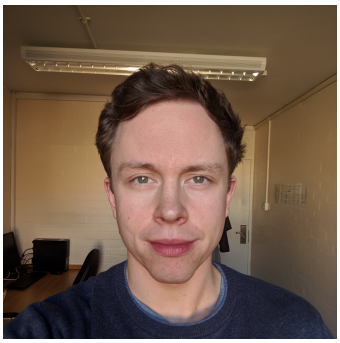
Yvonne Denier (Leuven) opened up with a very helpful and stimulating overview of the concept of 'taboo' that surrounds the limiting of health care: when setting thresholds that are related to cost-effectiveness, these will have implications of making trade-offs between individuals, illnesses, and treatments that are not easily discussed in the open. Partly, this is because the trade-offs force us to make the incommensurable commensurable. And partly, because any such trade-offs make us queasy. Her discussion was followed up by an instructive talk of Marcel Verweij (Wageningen) who investigated the moral import of these kinds of trade-offs against his practical experience at the Dutch National Health Care Institute (*Zorginstituut Nederland*) which determines and advises on which types of health care are included in the basic care package that everyone in the country is entitled to. In the work of this institute, decisions are made about the scope and contents of basic health care on the basis of cost effectiveness for health improvements. He explored the role of the concept of 'solidarity' in this context. Peter Kooreman (Tilburg) put these discussion in a global setting, showing that the costs per life saved differ greatly between countries, and also demonstrating that the effectiveness of prevention is often underestimated in discourses about health care. He investigated to what extent Nudge-type solutions should be put to use in the context of prevention.

Daniel Hausman (Wisconsin-Madison) as well as myself investigated the tradeoffs between considerations of fairness and goodness in making hard choices with regards to health. Considerations of fairness and goodness can come apart: consider the case of one donor kidney, and two individuals who need it to have their life saved. Ann would live for another 25 years, and Bob for another 20 years. Now, the best action is to give the kidney to Ann. But what about fairness? Both Ann and Bob have a claim to have their life saved. If fairness is equal satisfaction of claims, then giving the kidney to either one of them is unfair. In fact, the fairest action is to destroy the kidney, as then both their claims receive equal (zero) satisfaction. Broome has famously argued for holding a lottery between the two individuals, so that both of them get equal 'surrogate satisfaction' of a fair chance of winning the lottery, and one of them actually receives it, therefore also realizing an appropriate amount of good. This raises many issues about the interaction of goodness and fairness in the face of making limiting or rationing choices about health care. Indeed, it is easy to see how considerations of cost effectiveness, commensurability, solidarity, and taboos enter the mix, especially in complex real-world cases.

All participants found many more fruitful connections than those I mentioned – indeed, more than I can reasonably record here: the workshop was not only the presentation of ideas, but also the start of many more discussions and exchanges in the time to come. Which makes it as successful an event as any. A remarkable achievement of the organizers!

Conrad Heilmann
Erasmus Institute for Philosophy and Economics (EIPE)
Erasmus University Rotterdam

Evidence-Based Medicine



The big EBM news last month was that anti-depressants were declared to be effective. A network meta-analysis, the largest of its kind on these drugs (522 studies included with 116,477 participants) found that all major anti-depressants were more efficacious than placebo, and also which of those anti-depressants were the most efficacious when put head-to-head. Allegations of malign industry influence usually dog such trials, but the authors had access to numerous unpublished studies, and concluded that industry influence is not behind claims of efficacy. The study made the news headlines, to the degree that apparently a million more people should now be taking the drugs, and the BMJ also ran articles that spread the good word. This was backed up by clinical experts: Ilan Young, psychiatry professor at King's College London, told the BMJ: "Network meta-analyses are now widely accepted but depend on the data put in...[t]his study used a large amount of high quality data so it can be trusted." What can we make of these claims of quality and trust? I will offer some of my own thoughts about the quality of this evidence, but first I will put into perspective the size of the effect demonstrated in the study with a look at one of the critical responses to the study

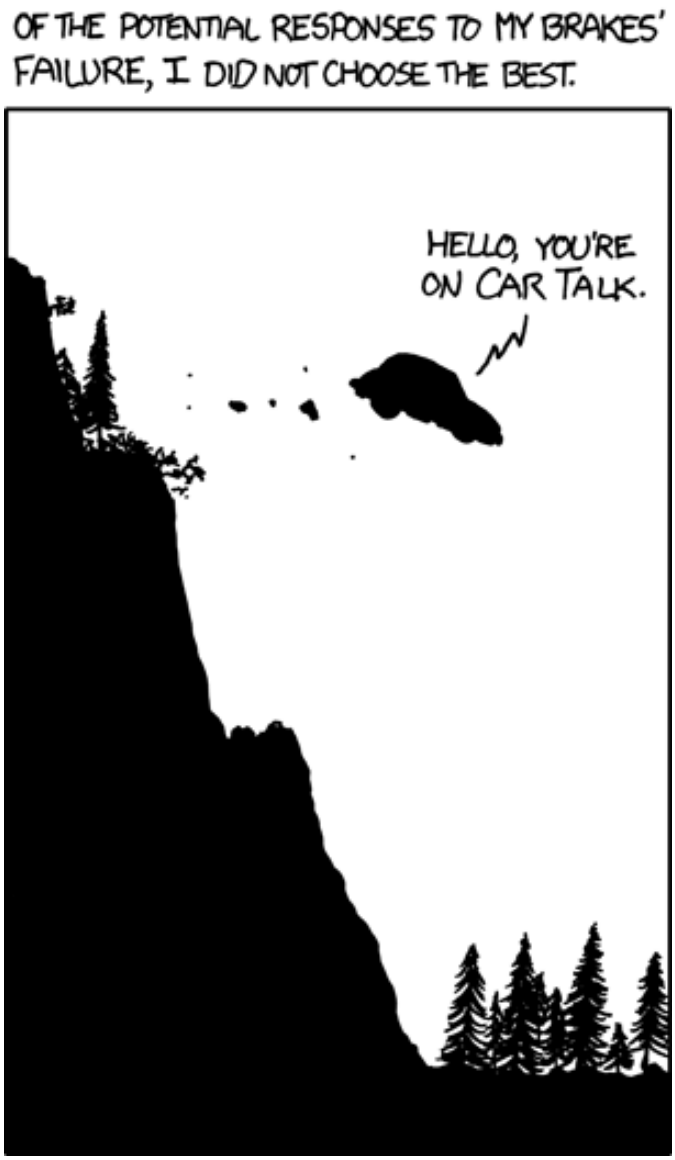
Using some plausible assumptions about the size of the placebo effect in the use of anti-depressants, a response in the BMJ concluded that given the study's average effect of an odds ratio of 1.6, this would mean only 10-12% more people in the treatment group would benefit compared with the placebo group. Further, of those in any group that responded favourably to the drug, 80% of those can be attributed to placebo. This shows how small the effect size is, but it is an effect size nonetheless. Another problem is that the study says nothing much about the harms of the drugs, both short and long term, or about the effectiveness of the drugs versus non-drug treatments like cognitive behavioural therapy. Crucially, it cannot tell us which antidepressant to prescribe for who. Weighing all of these practical considerations against such a small effect size makes guiding prescription difficult and this seems lost in the enthusiastic headlines.

What if we accept that clinical decision makers can handle nuance? Would this mean the claims of efficacy are warranted? Well this would depend on the quality of the evidence and this was assessed in two ways. For all drug v placebo trials, quality was assessed using 7 established criteria for rating risk of bias. 9% of the trials were rated as high risk of bias, 73% as moderate, and 18% as low. Just stating that there is a presence of risk of bias tells us nothing about how those biases will impact the effect size, but the mere fact of presence of biases should make us start to doubt the conclusions of efficacy. An odds ratio of 1 means 'no effect' - all it would take for the average anti-depressant to be considered no better than placebo is to have its estimate shifted downwards, or biased, by 0.6.

What about the head-to-head trials? Surely if some drugs have greater effects than others, then those better drugs are having some sort of effect? For these trials, to assess quality the authors rated the 'Certainty in the Evidence', using the GRADE system. Importantly, using GRADE means we can start to think about how the effect size is changed in the presence of biases. For GRADE, Certainty is a measure of the degree of confidence one has that the estimate of effect (usually the confidence interval) contains the true effect, and is also the same by definition and process of assessment as the 'Quality of Evidence'. If one's Certainty is lower than High-Certainty then there is more doubt than usual as to whether the confidence interval contains the true effect. We start to be more confident that the true effect actually lies outside the interval, often to a lower effect size than reported. Crucially, our confidence is effected by consideration of errors and biases made during the implementation and analysis of studies, thus making the link between bias and effect explicit. In the meta-analysis, all head-to-heads were rated either moderate, low or very-low Certainty. All were at least moderate, as on GRADEs criteria, the authors could not rule out 'publication bias' caused by industry influence. I will assume for now that this is not an issue due to the conclusions on industry influence made by the authors. Even allowing this, the majority of judgements of Certainty would be moderate or low. Given the small effect size detected in these trials, this may mean that most effect sizes are actually closer to 'no difference' (again, an odds ratio of 1). For example, Amitryptiline, the most effective overall against placebo and head-to-head, was moderate or low Certainty in 75% of trials. Further, only 3 of the 17 head-to-heads had an interval estimate that was entirely above 'no difference', and these three trials were all moderate to low Certainty trials. The true effect, on this judgement, could be in the 'no difference' region. This pattern holds for most average to good treatments. It is right that we can only trust a meta analysis due to the quality of its evidence. And the quality really doesnt look that good here.

To sum up, there are problems with recommending anti-depressants based on such small effect sizes given the acknowledged harms that come with their administration. Further, there is good reason to doubt that these small effect sizes are representative of the true effect size, even for the strongest drugs. The claims made from this study do not, as the headlines say, put the controversy to bed. Such headlines do not take account of the high degree of *Un*certainty present in the results of this trial.

Daniel Auker-Howlett
Philosophy, University of Kent



Events

April

JT: Just Theorising: Working Towards Responsible Methodologies, University of Sheffield, 9–10 April.
ETaGE: Epistemic Tools and Goods in Education, University of Pavia, 16 April.
MotM: Models of the Mind: Reasoning About Oneself and About Others, University of Edinburgh, 19 April.
FRaC: Formal Reasoning about Causation, Thessaloniki, Greece, 20 April.
CMoA: Computational Models of Argument, Warsaw, Poland, 20 April.

May

PMII: Perception, Mental Imagery and Inference, Ruhr University, Bochum, 14–15 May.
KBE: Knowledge, Belief, Evidence, University of Oxford, 21–23 May.
MMM: Modern Modeling Methods, University of Connecticut, 21–24 May.
E&U: Explanation and Understanding, Ghent University, 23–25 May.
ICAIBD: International Conference on Artificial Intelligence and Big Data, Chengdu, China, 26–28 May.

June

HaSE: Workshop on History and Scientific Explanation, KU Leuven, Belgium, 15–16 June.
RiPTW: Reasoning in a post-truth world: a look at dual-process models, Utrecht, the Netherlands, 20–21 June.
CMP: Computational Modeling in Philosophy, The Munich Center for Mathematical Philosophy, 22–23 June.
Logical Geometry and its Applications: Vichy, France, 25 June.

Courses and Programmes

Courses

LUCG: Logic, uncertainty and games, Como, 9–13 July.
SIPTA: 8th School on Imprecise Probabilities, Oviedo, 24–28 July.

Programmes

APhil: MA/PhD in Analytic Philosophy, University of Barcelona.
Master Programme: MA in Pure and Applied Logic, University of Barcelona.
Doctoral Programme in Philosophy: Language, Mind and Practice, Department of Philosophy, University of Zurich, Switzerland.
Doctoral Programme in Philosophy: Department of Philosophy, University of Milan, Italy.
LogiCS: Joint doctoral program on Logical Methods in Computer Science, TU Wien, TU Graz, and JKU Linz, Austria.
HPSM: MA in the History and Philosophy of Science and Medicine, Durham University.
Master Programme: in Statistics, University College Dublin.
LoPhiSC: Master in Logic, Philosophy of Science and Epistemology, Pantheon-Sorbonne University (Paris 1) and Paris-Sorbonne University (Paris 4).
Master Programme: in Artificial Intelligence, Radboud University Nijmegen, the Netherlands.
Master Programme: Philosophy and Economics, Institute of Philosophy, University of Bayreuth.
MA in Cognitive Science: School of Politics, International Studies and Philosophy, Queen's University Belfast.
MA in Logic and the Philosophy of Mathematics: Department of Philosophy, University of Bristol.
MA Programmes: in Philosophy of Science, University of Leeds.
MA in Logic and Philosophy of Science: Faculty of Philosophy, Philosophy of Science and Study of Religion, LMU Munich.
MA in Logic and Theory of Science: Department of Logic of the Eotvos Lorand University, Budapest, Hungary.
MA in Metaphysics, Language, and Mind: Department of Philosophy, University of Liverpool.
MA in Mind, Brain and Learning: Westminster Institute of Education, Oxford Brookes University.
MA in Philosophy: by research, Tilburg University.
MA in Philosophy, Science and Society: TiLPS, Tilburg University.
MA in Philosophy of Biological and Cognitive Sciences: Department of Philosophy, University of Bristol.
MA in Rhetoric: School of Journalism, Media and Communication, University of Central Lancashire.
MA programmes: in Philosophy of Language and Linguistics, and Philosophy of Mind and Psychology, University of Birmingham.
MRes in Methods and Practices of Philosophical Research: Northern Institute of Philosophy, University of Aberdeen.
MSc in Applied Statistics: Department of Economics, Mathematics and Statistics, Birkbeck, University of London.
MSc in Applied Statistics and Datamining: School of Mathematics and Statistics, University of St Andrews.
MSc in Artificial Intelligence: Faculty of Engineering, University of Leeds.

MSc in Cognitive & Decision Sciences: Psychology, University College London.
MSc in Cognitive Systems: Language, Learning, and Reasoning, University of Potsdam.
MSc in Cognitive Science: University of Osnabrück, Germany.
MSc in Cognitive Psychology/Neuropsychology: School of Psychology, University of Kent.
MSc in Logic: Institute for Logic, Language and Computation, University of Amsterdam.
MSc in Mind, Language & Embodied Cognition: School of Philosophy, Psychology and Language Sciences, University of Edinburgh.
MSc in Philosophy of Science, Technology and Society: University of Twente, The Netherlands.
MRes in Cognitive Science and Humanities: Language, Communication and Organization: Institute for Logic, Cognition,

Language, and Information, University of the Basque Country (Donostia San Sebastián).
Open Mind: International School of Advanced Studies in Cognitive Sciences, University of Bucharest.
Research Master in Philosophy and Economics: Erasmus University Rotterdam, The Netherlands.

Jobs and Studentships

Jobs

Post-doc: in Philosophy of Science, Ludwig Maximilian University of Munich, deadline 15 April.
Professorship: in Statistics and Data Science, the University of Western Australia, 23 April.
Post-doc: in New Epistemological Perspectives on Scientific Objectivity, University of Lyon, deadline 30 April.

Studentships

PhD position: in Machine Learning and Biomedicine, University of Edinburgh, deadline 6 April.
PhD position: in Statistics, the University of Kent, deadline 9 April.
PhD position: in Theoretical Philosophy, Stockholm University, deadline 16 April.
PhD position: in Decision Making, Delft University of Technology, deadline 23 April.
PhD position: in Computational Statistics, Delft University of Technology, deadline 1 May.
PhD position: in philosophy of science/ epistemology / philosophy of mind/cognitive science, Tilburg University, deadline 15 May.
3 PhD positions: in ethics of science/philosophy of science, two at Leibniz Universität Hannover, one at Bielefeld University, deadline 20 May.