MARIUSZ URBAŃSKI

Formal modeling of human reasoning: errors, limitations and Baconian bees

Mariusz Urbański

Adam Mickiewicz University,
ul. Wieniawskiego 1 61-712 Poznań, Poland.
E-mail: Mariusz.Urbanski@amu.edu.pl

Abstract: Since the end of the XX century we are witnessing a practical, or cognitive, turn in logic. Drawing on enormous achievements brought about by the mathematical turn that started more than a hundred years ago, logic now has come back to its Artistotelian roots as an instrument by which we come to know anything. The re-forged alliance between logic — now well equipped with sophisticated formal tools — and psychology results in more and more substantial developments in studies on human reasoning and problem solving. To reap the fruits of this alliance we need to be aware that it leads to a shift in focal points of interest of such studies as well as to expansion of their methodological repertoire. In this paper I argue that the practical turn in logic results in: (1) the concept of error becoming crucial for formal modeling of human reasoning processes, (2) prescriptive perspective, which takes into account human limitations in information processing, becoming the most interesting vantage point for such research and (3) triangulation of formal methods, quantitative approach and qualitative analyses becoming most effective methodology in formal modeling studies.

Keywords: practical turn in logic, formal modeling, error, reasoning

For citation: Urbański M. “Formal modeling of human reasoning: errors, limitations and Baconian bees”, *Logicheskie Issledovaniya / Logical Investigations*, 2020, Vol. 26, No. 2, pp. 106–115. DOI: 10.21146/2074-1472-2020-26-2-106-115

1. Where from, where to?

It has been a long journey for logic. It started with the Aristotelian idea of logic as an instrument (the ‘organon’) by means of which we come to know anything. Logic went down this path for a long, long time, finally reaching the psychologistic stance that “logic is the part of psychology that investigates relations between thinking and the reality” [Beneke, 1832, p. 12], that it is “the branch of inquiry [...] in which the act of the mind in reasoning is considered” [de Morgan, 1847, p. 26], that it searches for “the fundamental laws of those

operations of the mind by which reasoning is performed [and] some probable intimations concerning the nature and constitution of the human mind” [Boole, 1854, p. 1]. But shortly afterwards logic has fallen into antipsychologistic chasm in which Frege [Frege, 1884, p. 38] claimed that logic (and mathematics) is the most exact of all sciences, while psychology is imprecise and vague, and Husserl [Husserl, 1901, §§19–23] added that actual thinking is not driven by the laws of logic. Thus, the one has nothing to do with the other. As a result, logicians and psychologists started to behave “like men and women in the orthodox synagogue. Each group knows of the existence of other, but ignoring the fact is considered a proper form of behavior” [Macnamara, 1986, p. 1]. Logicians claimed that logic is not a science of reasoning. At most, one of the objects of its interests may be correctness of thinking, related to criteria based on objective — and therefore non-psychological — characteristics of linguistic expressions, such as their structure, or logical values. Psychologists, in turn, eagerly agreed that logic is not a science of reasoning, and added that it is completely useless in analysis of their correctness, since objective criteria set forth by logic are completely unrealistic and do not fit to how people really think. However, it soon turned out that the chasm is in fact a diamond mine and that mathematical turn in logic brought enormous developments in conceptual clarity, tools available as well as their sharpness and precision. Step by step – with Tarski’s formal semantics, Gödel limitations theorems, developments in philosophical logic, logical pragmatics, logic in AI and even proper logical analysis of fallacies – logical theories and systems started to look like a useful tool for real-life applications. Diamonds in the mine are just some hard objects and their true beauty demands sunlight to be appreciated. Thus, here we go again with the idea of logic as an instrument. With this practical [Gabbay and Woods, 2005], or cognitive [Urbański, 2011] turn logic acquires a new task of “systematically keeping track of changing representations of information” [van Benthem, 2008, p. 73] (and not only linguistic representations), and it becomes more and more capable of modeling actual cognitive activity of real life agents. The New Psychologism, or Cognitivism, claims that it is not the case that logic has nothing to do with how we think and also contests the claim that distinction between descriptive and normative account on analysis of reasoning is disjoint and exhaustive. Twelve years ago van Benthem [Ibid., p. 77] imagined “a world where a logician who has created a new logical system does two things instead of one: like now, submit to a logic conference, usually far abroad, but also: telephone the psychologist next door to see if some new nice experiment can be done”. Today we are even talking to each other face to face (see [Stenning and van Lambalgen, 2008; Zajenkowski et al., 2014; Żelechowska et al., 2020] for instructive examples).

2. What is the Precious?

But now: what the topic of these conversations should be? We, who are interested in formal modeling of human reasoning, in identifying regularities in human problem-solving which can be accounted for by logic — what exactly are we supposed to be focusing on? One obvious answer is, that we should be looking for norms against which quality of human reasoning may be evaluated. Establishing such norms allows for correctness of reasoning to be defined, allows for comparative assessment of both the process — reasoning — and the product — conclusions — to be carried out. As a result we would be able to discriminate between what's good and what's bad in human reasoning, what's safe and what's not from the point of view of knowledge development, what — ultimately — brings us closer to, well, the Truth. And this is a legitimate and well justified approach. However, I'd like to propose a different standpoint. If we are interested in looking through the window into reasoning processes and in learning about what humans really do while processing information, and if we are interested in how they could perform better, then our Holy Grail is the flipside of normative correctness: it is the error. It is the error which offers us an insight into how humans solve problems, how they process information, how they reason. If all the experimenters have access to in their studies are data — solutions to reasoning tasks — that are correct from a normative point of view, then all they know is that their subjects are probably well educated in formal methods, like logic, programming or so. More often than not the tasks themselves are not exemplary of real life problem-solving, so correct solutions do not tell much about real life problem-solving either. Analyses of errors, systematic errors in particular, allow to apply our very sophisticated formal tools to model actual reasoning processes.

Certainly, in order to identify something as an error you need to relate it to some concept of correctness. Thus, I am not claiming that we should abandon studies of formal logical systems: they are indispensable if we're going to step beyond idiographic studies of particular episodes of reasoning into the realm of modeling inter-subject regularities of information processing. What I am claiming is, that it is error in reasoning which lets us learn what these inter-subject regularities really are. So, it is more like a matter of where to start and how to proceed: top-down, implementing some standard of correctness, or bottom-up, identifying it with a little help of the data. I prefer the latter approach.

The concept of norm and normative correctness is somewhat ambiguous. Some time ago we abandoned the idea of deduction as the normative yardstick for evaluation of reasoning quality [Evans, 2012]. Now we have a plethora of different accounts on information processing which offer different takes on

what does it mean for a reasoning to be correct (or, at least, more correct than others). To some extent the same holds in case of the concept of error. However, in case of error the ambiguity stems not just from the fact that there are so many normative standards available against which a reasoner may sin. There are many substantially different modes of erring in reasoning. I shall identify just a couple of them. In that, I'll employ an abstract characteristic of the logical structure of reasoning, proposed by [Kurtz et al., 1999]:

$$y = F(x, k)$$

In this schema y represents the inferential product of the reasoning process (the conclusion). F stands for a summary of the set of computational tools used to manipulate, recombine or transform the input information. The initial available information (the premises) is represented by x . Finally, k stands for stored knowledge and experience, not identified explicitly as a part of x .

Now let's turn to the modes of erring in reasoning.

1. Your conclusion does not follow from your premises, which means that your F is faulty. This is probably the most obvious possible mode of error to be construed. It begs the question: in what sense of 'follow', exactly? Which concept of entailment is violated?
2. Your premises are not sufficiently justified: reliability of this particular x is dubious. This raises the issue of what does it mean to be sufficiently justified?
3. You misinterpreted the premises: your x is not exactly what you consider it to be. A good example here is deontological interpretation of the rule in the Wason Selection Task in the abstract setting [Stenning and van Lambalgen, 2008, see ch. 3], but with a caveat: what is a misinterpretation from the point of view of the eager experimenter need not to be such from the point of view of the subject.
4. You misconstrued the task: you applied F which is not best suited for what you were supposed to perform. It's a kind of exam pitfall, when e.g. the students asked to compare concepts A and B just give their definitions instead.
5. Your reasoning leads you to some kind of wilderness, or desert, or snake pit. It's again an F -related problem: we usually reason not just for the sheer fun of the act, but in a goal-directed manner, with some purpose guiding the process. This issue has a lot to do with Hintikka's [Hintikka, 2007] distinction of definitory rules (which are merely permissive, like in

deduction) vs strategic rules (which tell what to do in order increase one's chances of reaching the goal).

6. You are not able to gather information needed to even start the reasoning. Here F does not kick off with the available x , so to speak, or you lack appropriate k : this is the Holmesian 'I need data!'.
7. You ask for information which is not needed for your purposes: your x is not relevant to the task at hand (which is kind of the flipside of the previous one).
8. You're not able to process the amount of information needed in the time available. This is an F issue, but significantly going beyond the abstract account of [Kurtz et al., 1999]: our bounded resources enter the picture here, again related to the Hintikka's strategic perspective and also to the Peircean distinction of corrolarial vs theoremic reasoning [Levy, 1997]. This may be just a selection or an ability issue.
9. You don't know which tool to use, how to process your initial information: F issue again, more ability- than just selection-related.

3. What is the measure?

An error in reasoning occurs because an agent fails to meet certain criteria of correctness. What are they, then? What type of yardstick are we going to use in order to assess quality of reasoning? [Stanovich, 1999] introduced a useful distinction between three types of rules which may be used to this end; *a fortiori*, these are also three different perspectives in research on the topic (see also [Stenning and van Lambalgen, 2008, p. 6]):

- normative rules, that account for reasoning as it ideally should be;
- descriptive rules, that address reasoning as it is actually practiced;
- prescriptive rules, that result from taking into account our bounded rationality, i.e., computational limitations and storage limitations.

Logicians are probably not that much interested in descriptive rules (unlike psychologists, cognitive scientists or argumentation theorists), but along normative rules, prescriptive ones are of interests to them. We the humans have neither all the time in the world nor all the computational power needed to solve our problems by reasoning, thus we should be very much interested in answering the question: What is the best use to which humans may put their bounded reasoning and problem solving resources? Van Benthem captures this

idea with his remark that what is now of crucial importance to logic is not the static notion of correctness, but the dynamic one of correction and with the following claim:

Logic is of course not experimental, or even theoretical, psychology, and it approaches human reasoning with purposes of its own. And a logical theory is not useless if people do not quite behave according to it. But the boundary is delicate. And I think the following should be obvious: if logical theory were totally disjoint from actual reasoning, it would be no use at all, for whatever purpose! [van Benthem, 2008, p. 69]

But then another issue arises instantly: what exactly can we achieve with our formal models being applied to human reasoning processes? What types of regularities can we discover? Again, Stanovich [Stanovich, 1999, pp. 4–9] together with Stenning and van Lambalgen [Stenning and van Lambalgen, 2008, pp. 6–7] offer a useful frame to determine one’s position on this issue in terms of four possible stances on the relationship between reasoning and normative/prescriptive standards:

- Panglossian: Human reasoning competence and performance is actually normatively correct. What appears to be incorrect reasoning can be explained by such maneuvers as different task construal, a different interpretation of logical terms, etc.
- Apologist: Actual human performance follows prescriptive rules, but the latter are in general (and necessarily) subnormal, because of the heavy computational demands of normatively correct reasoning.
- Meliorist: Actual human reasoning falls short of prescriptive standards, which are themselves subnormal; there is therefore much room for improvement by suitable education.
- Eliminativist: Reasoning rarely happens in real life, and mainly in institutional contexts such as schools. By contrast, true rationality is adaptiveness: we have developed “fast and frugal algorithms” which allow us to take quick decisions which are optimal given constraints of time and energy.

At this point it should be obvious that wherever we start, from the top or from the bottom, at some point we cannot anymore ignore the concept of a standard, against which human performance in reasoning could be evaluated.

One trouble is that the Sun of classical logic is no longer the center of the logical universe. Even the Milky Way of logics employing a monotonic version of entailment is but a set of options to be chosen or not. With mathematicised logical systems developed so widely and deeply we have now a tool for every occasion: in fact one can probably model any way of processing information by using logical tools. So, are there any errors in reasoning — really? Isn't it that Panglossians are right, albeit not in their intended way: you may always find some norm according to which your reasoning will be correct, thus anything goes? And what about prescriptive standard becoming a kind of acting norm in lack of the proper one? One possible argument would be that it is a matter of selection of the contextually appropriate standard. But this is a fast-track slippery slope into a version of Aristotelian third man argument. Apparently, fundamental issues of the philosophy of science are inescapable in science. With the bottom-up approach I prefer to stay low, down-to-earth even, endorsing a position quite close to Haack's [Haack, 1993] foundherentism, imposing two kinds of constraints on each epistemic decision (as Haack does on rational belief): experiential anchoring and explanatory integration. Thus, I claim that an appropriate methodology for (mind you, interdisciplinary) research on modeling real human reasoning must involve a careful system of checks and balances of formal and computational methods, qualitative analyses and statistical quantitative approach. Such a triangulation brings about its own problems, but safeguards that reasonable amount of attention is paid to intertwined data of different character, as witnessed in, e.g., [Urbański et al., 2016; Urbański and Grzelak, 2019; Żelechowska et al., 2020].

Research so construed are located somewhere inbetween Apologist and Meliorist stances, probably closer to the latter. Occasionally we reach prescriptive standards in reasoning, but this requires some effort. Thus, more often than not we fall short of them, and there is plenty of room for appropriate logical/critical thinking education [Kisielewska et al., 2016].

4. A paradigm shift

In [Urbański and Klawiter, 2018, pp. 588–590] we have argued that there are three types of questions answers to which could be of interest in any scientific endeavour. The first one is substantialist question: What is X ? This is a question concerning the essence, or nature, or definition of a phenomenon or object under consideration. The second, functionalist question asks: What X is for? What is its function? Finally, the instrumentalist question concerns the use of X : how it is used? The focus on substantialist question is the key feature of what [Evans, 2012] identifies as the 'old' paradigm in the research on deduction. It starts with logical entailment as its constitutive property

and only eventually proceeds to noticing that deduction is sometimes used in episodes of real reasoning. The ‘new’ paradigm, characterized by a situated approach, starts with the identification of the contexts in which deduction is used. Within this new paradigm reasoning, judgement, decision making are viewed as involving similar processes. As a result, within the new paradigm deduction loses its privileged status as the ultimate normative yardstick for the evaluation of the correctness of reasoning and becomes more a strategic concept, “a form of reasoning that high-ability participants might engage in when suitably instructed and motivated to make deductive effort. [...] Rather than being a built-in function of the human mind, deductive reasoning can be seen as just one of many kinds of problem solving and formal thinking in which people of sufficient IQ can engage” [Evans, 2012, pp. 7–8]. *Mutatis mutandis*, a similar shift may be observed in case of other types of reasoning, although not in all the cases the analogy is exact. In case deduction or induction we do have precise answers to substantialist question. Thus, we may start with the definition of the type of reasoning in question, go to the data, if needed, and then go back, if necessary. It’s different in case of, for example, abduction. There is no satisfactory (that is, most of all, commonly accepted) answer to the substantialist question concerning abduction: we do not have a precise definition of this type of reasoning. What we have is just some general schema proposed by Peirce [Peirce, 1958, 5.189], that may be interpreted in a number of ways. In this case to start with the data is indispensable: it is an answer to the instrumentalist question which is of crucial importance, an answer to the functionalist question is a derivative and to the substantialist question — a nice bonus [Urbański and Klawiter, 2018, p. 591].

This or that way, reality of human reasoning processes becomes more and more important for logical investigations, be it as an empirical source of inspiration, a testbed or a figure to be at least partially represented in a formal system. Where do the models of reasoning processes come from, then? The possibility that they appear out of the blue is still legitimate: it is how they fit the data what matters, not their provenience. They may also be excavated from the data, as regularities in subjects’ reasoning or problem-solving reports. And finally, they may be cooked on our own, as it is often the case with ingenious applications of proof methods or modeling reasoning processes with Artificial Neural Networks.

Thus, if you want to account for human reasoning processes, subjected to limitations and errors, then put your logic to the proper Baconian use: in such an endeavour there is no place for Antmen or Spidermen, men of just experiment or men of just dogmas. Only the Beeman will be successful, the one who “gathers its material from the flowers of the garden and of the field,

but transforms and digests it by a power of its own” [Bacon, 1855, p. 78]. And if you’re serious about your formal models being about something *real*, you can’t escape some triangulation of formal methods, quantitative approach and qualitative studies, because practicality matters: “understanding interpretation sometimes leads to clarification of what subjects are trying to do, and that often turns out to be quite different than the experimenter assumes” [Stenning and van Lambalgen, 2008, p. 90].

Acknowledgements. Research reported in this paper were supported by the National Science Centre, Poland (DEC-2013/10/E/HS1/00172).

References

- Bacon, 1855 – Bacon, F. *The Novum Organon, Or a True Guide to the Interpretation of Nature: Book One*. Oxford: Oxford University Press, 1855. 375 pp.
- Beneke, 1832 – Beneke, F.E. *Lehrbuch der Logik als Kunstlehre des Denkens*. Berlin: Ernst Siegfried Mittler, Posen und Bromberg, 1832. 196 pp.
- Boole, 1854 – Boole, G. *An Investigation of the Laws of Thought*. London: Walton and Maberly, 1854. 424 pp.
- de Morgan, 1847 – de Morgan, A. *Formal Logic, or The Calculus of Inference, Necessary and Probable*. London: Taylor and Walton, 1847. 336 pp.
- Evans, 2012 – Evans, J.S.B.T. “Questions and challenges for the new psychology of reasoning”, *Thinking & Reasoning*, 2012, Vol. 18, No. 1, pp. 5–31.
- Frege, 1884 – Frege, G. *Die Grundlagen der Arithmetik: Eine logisch mathematische Untersuchung über den Begriff der Zahl*. Breslau: W.Koebner, 1884.
- Gabbay and Woods, 2005 – Gabbay, D.M. and Woods, J. *The Reach of Abduction. Insight and Trial*. Elsevier, 2005. 476 pp.
- Haack, 1993 – Haack, S. *Evidence and Inquiry: Towards Reconstruction in Epistemology*. Oxford: Blackwell, 1993. 272 pp.
- Hintikka, 2007 – Hintikka, J. “Abduction – Inference, Conjecture, or an Answer to a Question?”, in: *Socratic Epistemology. Explorations of Knowledge-Seeking by Questioning*. Cambridge: Cambridge University Press, 2007, pp. 38–60.
- Husserl, 1901 – Husserl, E. *Logische Untersuchungen*. Halle: Max Niemeyer Verlag, 1900/1901.
- Kisielewska et al., 2016 – Kisielewska, M., Urbański, M., and Paluszkiwicz, K. “Abduction in One Intelligence Test. Types of Reasoning Involved in Solving Raven’s Advanced Progressive Matrices”, in: *Model-Based Reasoning in Science and Technology*, Magnani, L. and Casadio, C. (eds.). Springer, 2016, pp. 419–435.
- Kurtz et al., 1999 – Kurtz, K.J., Gentner, D. and Gunn, V. “Reasoning”, in: *Cognitive Science, Handbook of Perception and Cognition*, B. M. Bly (ed.), 2d ed. Elsevier, 1999, pp. 145–200.
- Levy, 1997 – Levy, S.H. “Peirce’s theoremic/corollarial distinction and the interconnections between mathematic and logic”, in: *Studies in the Logic of Charles*

- Sanders Peirce*, Houser, N., Roberts, D.D., and Evra, J.V. (eds.). Bloomington, IN: Indiana University Press, 1997, pp. 85–110.
- Macnamara, 1986 – Macnamara, J. *A Border Dispute. The Place of Logic in Psychology*. Cambridge, MA: MIT Press, 1986.
- Peirce, 1958 – Peirce, C.S. *Collected Works*. C. Hartshorne, P. Weiss and A.W. Burks (eds.) Cambridge, MA: Harvard University Press, 1958.
- Stanovich, 1999 – Stanovich, K.E. *Who Is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum, 1999. 296 pp.
- Stenning and van Lambalgen, 2008 – Stenning, K. and van Lambalgen, M. *Human Reasoning and Cognitive Science*. Cambridge, MA: MIT Press, 2008. 424 pp.
- Urbański, 2011 – Urbański, M. “Logic and cognition: the faces of psychologism”, *Log. Log. Philos.*, 2011, Vol. (20), pp. 175–185.
- Urbański and Grzelak, 2019 – Urbański, M. and Grzelak, J. “On two simple models for one simple game: ‘guess who?’, Inferential Erotetic Logic, and situational semantics”, in: *Reasoning: Logic, Cognition, Games*, Urbański, M., Skura, T. and Łupkowski, P. (eds.). London: College Publications, 2019, pp. 1–18.
- Urbański and Klawiter, 2018 – Urbański, M. and Klawiter, A. “Abduction: some conceptual issues”, *Log. Log. Philos.*, 2018, Vol. (27), pp. 583–597.
- Urbański et al., 2016 – Urbański, M., Żyluk, N., Paluszkiewicz, K., and Urbańska, J. “A Formal Model of Erotetic Reasoning in Solving Some what Ill-Defined Problems”, in: *Argumentation and Reasoned Action Proceedings of the 1st European Conference on Argumentation, Lisbon 2015*, Mohammed, D. and Lewiński, M. (eds.). London: College Publications, 2016, pp. 973–983.
- van Benthem, 2008 – van Benthem, J. “Logic and Reasoning: do the facts matter?”, *Studia Logica*, 2008, Vol. 88, No. 1, pp. 67–84.
- Zajenkowski et al., 2014 – Zajenkowski, M., Garraffa, M., and Szymanik, J. “Working memory mechanism in proportional quantifier verification”, *Journal of Psycholinguistic Research*, 2014, Vol. 43, Is. 6, pp. 839–853.
- Żelechowska et al., 2020 – Żelechowska, D., Żyluk, N., and Urbański, M. “Find Out: a new method to study abductive reasoning in empirical research”, *Int. J. of Qualitative Methods*, 2020, Vol. 19, pp. 1–11.