Are Humans Rational?

Fall 2019 M R 4–5:50 DCC 337

Selmer Bringsjord
Rensselaer AI & Reasoning (RAIR) Lab
http://www.rpi.edu/∼brings
Department of Cognitive Science
Department of Computer Science
Lally School of Management
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA
Office Hours: SB: W 10–12 CA 3rd flr; & by appointment
Selmer.Bringsjord@gmail.com

ver 1107190830NY

Contents

1 General Orientation

1.1 Driving Thesis

The Aristotelian dictum that we are *rational animals* is under severe attack these days. In fact, the previous sentence may seriously understate the situation: the dictum is perhaps outright rejected by many, if not most. From psychologists of reasoning and decision-making to behavioral economists to the "new atheists" (all groups whose message we will consider in this class), the onslaught is firmly underway, and fierce. Yet this course revolves around a defense of the proposition that Aristotle, modernized along Leibnizian, Piagetian, and Bringsjordian $\times 2$ lines, is right. This proposition, put a bit more precisely, is:

> $\mathcal{R}$ Humans, at least neurobiologically normal ones, are fundamentally rational, where rationality is constituted by certain logico-mathematically based reasoning and decision-making in response to real-world stimuli, including stimuli given in the form of focused tests; but mere animals are *not* fundamentally rational, since, *contra* Darwin,[1] their minds are provably, fundamentally, qualitatively inferior to the human mind. As to whether computing machines/robots are fundamentally rational, the answer is also "No." For starters, if x can't read, write, and create, x can't be (presently[2]) rational; neither computing machines/robots nor non-human animals can read nor write nor create; ergo, they aren't fundamentally rational for this reason alone. But the news for non-human animals and computing machines/robots gets much worse, for they have not the slightest chance when they are measured against $\mathcal{H}$.

And here's $\mathcal{H}$ set out provisionally:

> $\mathcal{H}$ Humans have the ability to gain knowledge by reasoning (e.g., deductively) quantificationally and recursively over abstract concepts, including abstract concepts of a highly expressive, including infinitary, nature, expressed in arbitrarily complex natural and formal languages.

1.2 Rapid Example

For a rapid example[3] of some of the stimuli to which $\mathcal{R}$ refers:

Amtrak-to-Princeton J-L Problem

Suppose that the following two statements are true:

(1) Everyone likes anyone who likes someone.

(2) Abigail likes Bruno.

Does it follow deductively that everyone likes Bruno? Prove that your answer is right!

1.3 Notice: 'Fundamentally'

Notice that the adverb 'fundamentally' is used repeatedly in $\mathcal{R}$. This means, among other things, that humans are *potentially* rational. What humans need in order to reason and make decisions in the relevant ways, we (i.e., S & A) further claim, is sustained study of the relevant logic and mathematics, and an ability to use what one has studied in order to reason and decide correctly in response to the aforementioned stimuli. In the course of our defense, we're going to supply at least some of the relevant logic and mathematics to you. Hence, as you receive and judge our case, we believe that you will move some distance from being merely fundamentally rational to being *presently* rational. We also believe

[1] In his egregiously illogical *Descent of Man*.

[2] As opposed to potentially rational. The important distinction between *presently rational* vs. *potentially rational* will be taken up in class discussion early in the semester.

[3] Provided long ago on Amtrak to Selmer by Professor Yingrui Yang, who relayed it from Professor Johnson-Laird.

it's fair to say that the primary purpose of an undergraduate education is to markedly increase the level of reasoning and decision-making power that constitutes being presently rational.

1.4 A Disclaimer!

Please note that no collaborators or guest lecturers other than A Bringsjord, should be assumed to have affirmed anything like the claims $\mathcal{R}$ and $\mathcal{H}$ issued above. This thus applies specifically to TA Can Mekik, as well as to any RAIR-Lab researchers who may present or demo in class or by pre-recorded video (see §1.6). As to what these thinkers hold in connection with $\mathcal{R}$ and $\mathcal{H}$, that is an open question. You are free to inquire of them.

1.5 Context: A Research University

You have wisely decided to attend a technical *research* university, with a faculty-led mission to create new knowledge and technology in collaboration with students. RPI is the oldest such place in the English-speaking world; it may know a thing or two about this mission. The mission drives those who teach you in this class. The last thing we want to do is simply convey to you how *others* answer the driving question that gives this class its name.[4] As should be obvious by now, we think we have correct answers to the driving question, and are working hard to explain them, specify them formally, and disseminate them. We'll tell you objectively what other thinkers say, but we're going to tell you that, at least for the most part, they're wrong. You can judge whether our arguments for such negative dianoses are sound or not. And you should start to develop your own individual answer, which may well be different than ours. You should seek to defend your answer, and will indeed by asked to do so in this class. For purposes of evaluating your performance, it matters not a whit what your positions are; what matters is your understanding of the technical material presented, and the quality of your reasoning given in defense of your positions on the topics covered.

1.6 Graduate Teaching Assistants; Further Help

The TA for this course is: Can Mekik; email address: `mekikc@rpi.edu`. Can will hold office hours in the Cognitive Architecture Lab (41 9th St, Floor 1; aka EMPAC Annex; requires card access; ring bell/knock on front window to get in), on Thurdays from 1pm to 3pm. Additional assistance will be provided by RAIR-Lab researcher and PhD student Mike Giancola, who among other things will be giving some guest lectures in his areas of expertise in AI. Please note again §1.4.

2 Prerequisites

There are no formal prerequisites. However, this course covers parts of such things as formal deductive logic, formal probabilistic logic, basic probability theory, game theory, etc. This implies that — for want of a better phrase — students are expected to have a degree of "mathematical maturity." At RPI, this expectation is quite reasonable.[5]

[4] Calls by various people in the U.S. to supress/minimize the research role of faculty is self-refuting, at least within the setting of a research university, because since all fields advance by way of cutting-edge research and innovation taking place at such institutions, and since these advances and the fruit produced by them must be taught, any supression/minimization of the research role of faculty at research universities is by definition to compromise teaching itself. In short, the only antidote to teaching old and stale material is to infuse teaching with the creation of new material, preferably by the teacher in question, in collaboration with his/her students.

[5] To be a bit more specific, the logico-mathematics alluded to in claim $\mathcal{R}$ can be partitioned into three general areas: analysis and continuous mathematics (A1); deductive formalisms, systems, and techniques (A2); and inductive/statistical/probabilistic formalisms, systems, and techniques (A3). Because of the nature of RPI's requirements for a BS, A1 is generally already covered in other classes (e.g. differential and integral calculus). The emphasis in the present class is on (introductory elements of) areas A2 and A3. Please be aware that some available coverage of the formal material

3 Texts/Readings

In-class lecture, discussion, and debate deliver the crucial content; indeed, the primary content to be learned is obtained in class. (Assuming that things go according to plan, all lectures will be recorded, and will be available for review to all students.) Attendance is required and will be taken, and note-taking is key. Sometimes slides will be distributed by posting on the course web page. Most readings will be electronic, and either distributed by email, or can be obtained by url (often the slide decks will have urls). As a first example, students should read (Baker 2013) asap, since it (we claim) represents a stark example of an implicit denial of $\mathcal{R}$ and $\mathcal{H}$. As to books, it's required that students purchase and read Kahneman's (2013) *Thinking, Fast and Slow*. S Bringsjord will draw heavily from his forthcoming *Gödel's Great Theorems* (forthcoming) in five class meetings (see §4). The book *Humans 3.0: The Updgrading of the Species*, by P. Nowak, is recommended, and is available in the Rensselaer Bookstore (and also of course e.g. Amazon).

There is a second recommended book, available online, from which S Bringsjord will sometimes draw. For anyone serious about the study of rationality at the human level, the book in question is an absolute must read. The book in question is Robert Nozick's (1993) *The Nature of Rationality*. It's available as a pdf online at here. Of course, in this day and age, there's a lot more available online that's relevant to the present course, and as we proceed, links will be provided. In the schedule which follows, some links to readings are provided.

associated with systematic coverage of rationality is (irrationally) restricted to only formalisms commonly associated with micro-economics. An example is the book *Rational Choice* (2010) by Itzhak Gilboa — a nice book, certainly, and one I will refer to later, but a very narrow one in the context of our class here at RPI.

4 Schedule

Inevitably, what is actually covered, meeting by meeting, will vary a bit from what follows. The definitive record of what happens, and of content presented, and therefore of what students are responsible for on tests and papers, will be on the course web page, which includes a meeting-by-meeting repository of slide decks as the course progresses.

4.1 Setting the Stage

- **Aug 29**: General Orientation, Logistics, Mechanics. The syllabus is reviewed in detail. It's made clear to students that there is a very definite position (viz., $\mathcal{R}$ and $\mathcal{H}$) advocated in the class, and that content from the formal sciences will be presented at a fast pace. Students who, upon learning about the nature of *Are Humans Rational?* in this meeting, find that it's not their cup of tea, are encouraged to make a change in their schedule before next class.
- **Sep 2**: No Class (Labor Day)
- **Sep 3**: Main Claims $\mathcal{R}$ and $\mathcal{H}$ Presented and Initial Discussion **Stage I**, via Overarching Picture. Humans, despite recent claims to the contrary, are rational; more specifically, $\mathcal{R}$, which is the main claim. Rationality consists in cognition that conforms to relevant logic & mathematics, in the face of tests. Piaget was fundamentally correct that humans are fundamentally logical. The methods and anti-$\mathcal{R}$ claims of "disparaging" psychology of reasoning and decision-making are irrational and should be rejected. Corollary: Claims, such as N Baker's, that algebra should be optional, with all of mathematics, if heeded, would doom people to a pre-rational phase — which obviously would be a very bad thing. It would also doom them to disemployment by machines. If there's time, some consideration will be given to the view of E Sosa, which, *contra* the view expressed and defended by S Bringsjord, is that rationality shouldn't be defined via reference to formal logic, probability, etc.[6]

4.2 The Attack from Failures of Deductive Reasoning

- **Sep 5**: Now we move to Stage II in the presentation of the overarching picture, and defense and discussion thereof. Review and Expansion of "Main Claims Presented." Then: The Original, Classic Shots at Piaget (from Wason, Johnson-Laird, etc.). This includes the Wason Selection Task and the THOG Problem, e.g. We also examine more recent, harder problems (J-L's king-ace problem). The class ends with a refutation of the argument against $\mathcal{R}$ based upon human performance on the problems in question.
- **Sep 9**: Analysis of Shots @ Piaget (& @ Aristotle too), $\mathcal{R}$ from Previous Class. The analysis serves to begin our learning of the propositional calculus. The analysis is preceeded by a brief history of the deductive formalisms (including more prominently the propositional calculus) featured in $\mathcal{R}$.
- **Sep 12**: "Cognitive" Deductive Shots Considered. Shots. These shots too are rendered impotent, and are used to advance our understanding of some of the formalisms necessary to be rational. Class includes coverage of Floridi's poison-pill test of self-consciousness in AI/robots. Relevant background reading includes this paper, and this YouTube video. (The paper is not required reading, and is somewhat technical in places.)

4.3 The Attack from Failures of Probabilistic Reasoning

- **Sep 16**: Probabilistic Logic/Entailment. First a brief presentation by Mike Giancola on the role of the emotions in rational thought, according to John Pollock. Then we proceed to: Kolmogorovian axioms sets "declarativized." Normative correctness characterized with help from the basic Venn-Diagram approach, in so-called 'probability logic.' Also presentation, in miniature, of uncertainty calculus to be used later in conjunction with Lottery Paradox and St Petersburg Paradox. Does the Monty Hall Problem show that humans are irrational?

4.4 Refuting Kahneman

In the next two classes, we deflate two of Nobelist (economics) Kahneman's main attacks on rationality.

- **Sep 19**: Linda, Heuristics, & Logic. Reading from Kahneman must be studied beforehand. We present a new theory of "narratological" reasoning that explains why so many humans fail to correctly solve Linda-style probability problems.
- **Sep 23**: Overconfidence in "Professional" Investors; the Efficient Market Hypothesis. Reading from Kahneman must be studied beforehand. Bad Choices, Framing Effects, Prospect Theory. The case of Jim Simons is considered. Also considered is the case of activist investors, whose basic *modus operandi*

[6] Unfortunately, the relevant paper by Sosa (1999), entitled none other than "Are Humans Rational?", is hard to find online.

is defended as supremely rational (though certainly not invariably successful).

- **Sep 26**: **Test #1**: Basic Machinery of Rationality (from areas A2 & A3; see footnote 5.)

4.5 The Meaning of Life

- **Sep 30**: Pain, Pleasure, Utility & the Meaning of Life. Thagard, Kahneman, Nozick. Defense of Nozick's (1981) argument in Chapter 6 of his *Philosophical Investigations*.[7] Objections to Thagard's (2012) position and argument in *The Brain and the Meaning of Life* (a refined proper subset of those given in Bringsjord 2014). The "Experience Machine" and *The Matrix* turn out to be very relevant, and are discussed.

4.6 A Rational Treatment of AI and Machine Learning

- **Oct 3**: The Singularity. The claim that machines with superhuman intelligence will before long suddenly arrive is debunked. Evaluation of the original argument for The Singularity by Good. CRA & irrational fearlessness. AlphaGo, Schmalfa Go. How hard are these games — compared to other things humans do? We discuss the fundamental either-or of Gödel on comparing the power of our minds with computing machines, and here draw on the final chapter of (Bringsjord forthcoming).
- **Oct 7***: The Future of AI. There will be a sober assessment of today's "machine learning" techniques and technology, including "Deep Learning." Are machines really and truly learning? Or is there just a lot of hype? Bringsjord argues that the latter holds. For reading, see this paper.
- **Oct 10**:* The MiniMaxularity, & Human Disemployment. Fears about The Singularity are irrational, but if we calmly focus on what we *know* is coming from AI, there's reason to be concerned — despite what sanguine economists tell us. What should we then do? The distinctive (Descartes was right!) human power for: general problem-solving, robust language, creativity. Background reading is: (Bringsjord & Bringsjord 2017). There will also be coverage of Arrow's Impossibility Theorem, democracy, and rational moves to be made at the government level in light of the MiniMaxularity.

[7] See also for the argument Chaps. 15 & 16 in (Nozick 1989).

- **Oct 14**: No class (Columbus Day).
- **Oct 17**:* Making Morally X Machines (and Robots). We here consider a proposal for a rational, measured response to the concern on the part of the likes of Musk, Hawking, and Bostrom that autonomous machines pose an "existential" threat. The response is to engineer machines and robots that are ethically correct! This class provides an overview of the "Making Morally Competent Robots" MURI project at RPI, sponsored by the Office of Naval Research.
- **Oct 21**: **Test #2**: short essay questions, some of which call for issuing and defending a non-trivial claim.

4.7 The Paradoxes

- **Oct 24**:* Why study paradoxes? (In a word, because they focus the mind in the direction of careful, rigorous, *rational* thought.) We begin with The Liar & Russell's Bogus Barber. But then we shift to the The Knowability Paradox, and see whether a robot can navigate its turbulent waters. Mike Giancola will present on the dangers of inconsistency in some life-and-death situation, and on the desired capability of an AI to detect and manage inconsistency.
- **Oct 28**: Newcomb's Problem (original and Stalnaker's cancer variation). What's rational to do and/or decide?
- **Oct 31**:* The Lottery Paradox — or how a lottery almost led to a divorce for Selmer. This paradox is solved with RAIR-Lab technology! One solution provided by PhD student Mike Giancola — with some inspiration & help from the late John Pollock; a different solution, albeit in the same spirit, provided by Selmer Bringsjord. Can the St. Petersburg Paradox be solved as well? Yes!
- **Nov 4**:** The Paradoxes of Time Travel (grandfather paradox and looping painter). A key component of the discussion is Gödel's gift to Einstein: a proof that time travel (backwards) is possible. This class meeting in significant part draws from the chapter "Gödel's Time Travel Theorem," in *Gödel's Great Theorems*, by S Bringsjord, forthcoming from Oxford University Press.

4.8 Darwin's Dumb Ideas

- **Nov 7**: *Descent* into Error & Veneration of the Dog. Wallace's argument. Why Darwin would've flunked Logic 101 in light of his reasoning about human vs. non-human reasoning in his *Descent of Man* (Darwin 1997). Note: This isn't *Origin*! That's different, and a masterpiece.

- **Nov 11**: The great divide between human vs. non-human reasoning and problem-solving. This class includes an analysis, formalization, and discussion of the Penn et al. (2008) *BBS* paper "Darwin's Mistake," available at here. We also consider whether fish are so smart that maybe we should adjust our behavior toward them. The book *What a Fish Knows* (Balcombe 2016) is evaluated.

4.9 Animals & AIs? Can They Talk?

- **Nov 14**: The Marvel That is Human Linguistic Ability. Analysis and discussion of linguistic divide between humans vs. nonhuman animals (and we refer again here to PHP's paper), and humans vs. contemporary AIs. Chimps can't talk, and so Chomsky was right — and listening to him could've saved a lot of money. With the advent of Siri, Alexa, various chatterbots, etc., is there now reason to think that machines are closing the gap with us in the area of natural-language communication? Does AI translation technology likewise show that the gap is closing?
- **Nov 18**: The Astonishing Hardness of First Language Acquisition. Can we by the way apply this to Julian Jaynes's theory of how self-consciousness arose in our species, a theory which b.t.w. figures prominently in the HBO series *Westworld*?

4.10 Attack on Rationality From "New" Atheism

- **Nov 21**:** Rash Russell; Hapless Harris; Perspicacious Pascal. Breaking Dennett's *Breaking the Spell*. Focus includes Dennett's argument from inconsistency of the union of religious claims. Is religious belief irrational? Bringsjord presents and defends his new version of Pascal's Wager.

4.11 Steeple of Rationalistic Genius: Gödel

The next three classes each match up with a chapter from *Gödel's Great Theorems* by S Bringsjord, forthcoming from Oxford University Press.

- **Nov 25**: Gödel's Completeness Theorem and the Long Island Rail Road. Excerpted from *Gödel's Great Theorems*. (We will already have visited Gödel's Time Travel Theorem, recall; see **Nov 5** class.) This is an optional class.
- **Nov 28**: No Class (Thanksgiving Recess)
- **Dec 2**: Return of The Liar: Gödel's Incompleteness Theorem. Excerpted from *Gödel's Great Theorems*. We begin by taking stock of Hilbert's 1920 lecture, in which he set out the great problems for future mathematicians to tackle.
- **Dec 5**: The Continuum Hypothesis, and Gödel's Greatest Theorem. Excerpted from *Gödel's Great Theorems*.
- **Dec 9**: **Test #3**: short essay questions, *each* of which calls for issuing and defending a non-trivial claim.

5 Grading

Test #1: 15%. Test #2: 20%. Test #3: 25%. Paper: 20%. (The logistics of the paper will be explained later. It will be done in two submissions. The first submission will receive feedback and a tentative grade. A final version will be submitted that takes account of the feedback. The paper must be a direct attack on a claim and supporting argument of S Bringsjord's.) Class Participation: 20%. Test #1 will have few to no essays, as it will be focused on technical material. Test #2 will have some combination of multiple-choice, short-answer, and short- and medium-sized essay questions. Test #3, which is not cumulative, will have only essay questions. In the case of essays in which you must articulate and defend a determinate, non-trivial claim, you will need to give an argument for your claim, and at least one serious objection must be presented and rebutted. Hard-working students with the aptitude of those admitted to RPI have it within their power to receive an A in this course. With respect to the prior content in the present paragraph: One or more tests, and the paper itself, may be dropped if only one TA is provided for the course, an issue not settled at the moment.

6 Some Learning Outcomes

There are three desired outcomes.

O1 Students will understand the covered arguments against the theses $\mathcal{R}$ and $\mathcal{H}$.

O2 Students will understand the main covered Bringsjordian arguments and counter-arguments in favor of the thesis that (i) humans are fundamentally rational (= in favor of $\mathcal{R}$), while (ii) non-human animals and computing machines/robots aren't.

O3 Students will understand, to a significant degree, the relevant logico-mathematical terrain on which which debates over the driving question take place (e.g., propositional calculus, first-order logic, basic modal logic, probability logic, game theory, decision theory, etc.). See again footnote 5.

Q4 Students will gain appreciable skill at articulating and defending their claims by way of formidable arguments, verbally and in written form.

7 Academic Honesty

Student-teacher relationships are built on mutual respect and trust. Students must be able to trust that their teachers have made responsible decisions about the structure and content of the course, and that they're conscientiously making their best effort to help students learn. Teachers must be able to trust that students do their work conscientiously and honestly, making their best effort to learn. Acts that violate this mutual respect and trust undermine the educational process; they counteract and contradict our very reason for being at Rensselaer and will not be tolerated. Any student who engages in any form of academic dishonesty will receive an F in this course and will be reported to the Dean of Students for further disciplinary action. (The *Rensselaer Handbook* defines various forms of Academic Dishonesty and procedures for responding to them. All of these forms are violations of trust between students and teachers. Please familiarize yourself with this portion of the handbook.)

References

Baker, N. (2013), 'Wrong Answer: The Case Against Algebra II', *Harper's Magazine* pp. 31–38.
URL: *https://harpers.org/archive/2013/09/wrong-answer*

Balcombe, J. (2016), *What a Fish Knows: The Inner Lives of Our Underwater Cousins*, Scientific American / Farrar, Straus and Giroux, New York, NY.

Bringsjord, S. (2014), 'Review of P. Thagard's *The Brain and the Meaning of Life*', *Religion & Theology* **21**, 421–425.
URL: *http://kryten.mm.rpi.edu/SBringsjord_review_PThagard_TBTMOL.pdf*

Bringsjord, S. (forthcoming), *Gödel's Great Theorems*, Oxford University Press, Oxford, UK.

Bringsjord, S. & Bringsjord, A. (2017), The Singularity Business: Toward a Realistic, Fine-grained Economics for an AI-Infused World, *in* T. Powers, ed., 'Philosophy and Computing: Essays in Epistemology, Philosophy of Mind, Logic, and Ethics', Springer, Cham, Switzerland, pp. 99–119. The URL here is to a preprint only. This book s Volume 128 in Springer's *Philosophical Studies Series*, which is edited by Luciano Floridi and Mariarosaria Taddeo.
URL: *http://kryten.mm.rpi.edu/SBringsjord_ABringsjord_SingularityBiz_0915151500.pdf*

Darwin, C. (1997), *The Descent of Man*, Prometheus, Amherst, NY. A reprint edition. The book was first published in 1871.

Gilboa, I. (2010), *Rational Choice*, MIT Press, Cambridge, MA.

Kahneman, D. (2013), *Thinking, Fast and Slow*, Farrar, Straus, and Giroux, New York, NY.

Nozick, R. (1981), *Philosophical Investigations*, Belknap Press of Harvard University Press, Cambridge, MA.

Nozick, R. (1989), *The Examined Life*, Simon & Schuster, New York, NY.

Nozick, R. (1993), *The Nature of Rationality*, Princeton University Press, Princeton, NJ.

Penn, D., Holyoak, K. & Povinelli, D. (2008), 'Darwin's Mistake: Explaining the Discontinuity Between Human and Nonhuman Minds', *Behavioral and Brain Sciences* **31**, 109–178.

Sosa, E. (1999), Are Humans Rational?, *in* K. Korta, E. Sosa & X. Arrazola, eds, 'Cognition, Agency and Rationality', Kluwer, Dordrecht, The Netherlands, pp. 1–8. This book is the Proceedings of the Fifth International Colloquium on Cognitive Science.

Thagard, P. (2012), *The Brain and the Meaning of Life*, Princeton University Press, Princeton, NJ.