On *Breaking the Spell of Irrationality*

Selmer Bringsjord

*Are Humans Rational?*

11/23/15

RPI
Some Logistics
Some Logistics

• Dec 3 class: “AI as (Profound) Art”

• Pls sign up to attend, if desired, @

• https://webforms.rpi.edu/TCFA

• Mod: Only one Gödelian “Steeple of Excellence.”
Some Logistics

- Dec 3 class: “AI as (Profound) Art”
- Pls sign up to attend, if desired, @
  - https://webforms.rpi.edu/TCFA
- Mod: Only one Gödelian “Steeple of Excellence.”
- Am working on the “superset” questions for T3. Should be done & posted by Monday 11/30—but might be earlier.
Some Logistics

• Dec 3 class: “AI as (Profound) Art”

• Pls sign up to attend, if desired, @
  • https://webforms.rpi.edu/TCFA

• Mod: Only one Gödelian “Steeple of Excellence.”

• Am working on the “superset” questions for T3. Should be done & posted by Monday 11/30—but might be earlier.

• Grades for T1B: See me for last 5 mins of class today.
Some Logistics

• Dec 3 class: “AI as (Profound) Art”

• Pls sign up to attend, if desired, @ https://webforms.rpi.edu/TCFA

• Mod: Only one Gödelian “Steeple of Excellence.”

• Am working on the “superset” questions for T3. Should be done & posted by Monday 11/30—but might be earlier.

• Grades for T1B: See me for last 5 mins of class today.

• Have a great Thanksgiving holiday!
The Book

Once Broken, Religious People are Freed to be Truly Rational
Here’s how it works:
Here’s how it works:

• Theists and atheists share an affirmation of, and both in fact use, a common thing: *thinking tools* (= “cultural software”) that cut(s) across all human beings.
Here’s how it works:

• Theists and atheists share an affirmation of, and both in fact use, a common thing: thinking tools (= “cultural software”) that cut(s) across all human beings.

• Human beings, blessed as they are with a capacity for meta-reasoning and meta-representations and meta-representational capacity (recall ‘recursion’ and ‘hierarchical reasoning’ from PHR & our discussion of their BBS paper), can be brought to a realization that thinking tools, suitably deployed, entails the truth of atheism.
Here’s how it works:

- Theists and atheists share an affirmation of, and both in fact use, a common thing: *thinking tools* (= “cultural software”) that cut(s) across all human beings.

- Human beings, blessed as they are with a capacity for meta-reasoning and meta-representations and meta-representational capacity (recall ‘recursion’ and ‘hierarchical reasoning’ from PHR & our discussion of their BBS paper), can be brought to a realization that thinking tools, suitably deployed, entails the truth of atheism.

- So, deploy these tools and join the enlightened community of atheists!
the stable middle ground that Balkin provides: an open-minded ("ambivalent") stance that permits a rational dialogue to engage the issues between people, no matter how radically different their cultural backgrounds. We can engage in this conversation with some reasonable hope of resolution that isn't simply a matter of one culture overwhelming the other by brute force. We cannot expect, Balkin argues, to persuade others if we leave no room and opportunity for them to persuade us. Success does depend on the participants' sharing, and knowing that they share, two transcendent values of truth and justice. What this means is only that both parties accept that these values are inescapably presupposed by human projects that we all participate in, simply by being alive: the projects of staying alive, and staying secure. Nothing more parochial need be assumed, and even "Martians" should be able to agree on this.

The idea of a transcendent value is rather like the idea of a perfectly straight line—not achievable in practice, but readily comprehended as an ideal that can be approximated even if it can't be fully articulated. At first this may look like a dubious dodge—an ideal that we all somehow accept even if nobody can say what it is! But in fact, just such ideals are accepted and inescapable even in the most rigorous and formalistic of investigations. Consider the ideal of rationality itself. When logicians disagree about whether classical logic is to be preferred to intuitionistic logic, for instance, they have to have in mind a prior standard of rationality, by appeal to which one logic could be seen (by all) as better than another, and they have to presume that they share this ideal, but they don't have to be able to formulate this standard explicitly—that's what they're working on. And in just the same spirit, people with radically different ideas about which
the stable middle ground that Balkin provides: an open-minded ("ambivalent") stance that permits a rational dialogue to engage the issues between people, no matter how radically different their cultural backgrounds. We can engage in this conversation with some reasonable hope of resolution that isn’t simply a matter of one culture overwhelming the other by brute force. We cannot expect, Balkin argues, to persuade others if we leave no room and opportunity for them to persuade us. Success does depend on the participants’ sharing, and knowing that they share, two transcendent values of truth and justice. What this means is only that both parties accept that these values are inescapably presupposed by human projects that we all participate in, simply by being alive: the projects of staying alive, and staying secure. Nothing more parochial need be assumed, and even “Martians” should be able to agree on this.

The idea of a transcendent value is rather like the idea of a perfectly straight line—not achievable in practice, but readily comprehended as an ideal that can be approximated even if it can’t be fully articulated. At first this may look like a dubious dodge—an ideal that we all somehow accept even if nobody can say what it is! But in fact, just such ideals are accepted and inescapable even in the most rigorous and formalistic of investigations. Consider the ideal of rationality itself. When logicians disagree about whether classical logic is to be preferred to intuitionistic logic, for instance, they have to have in mind a prior standard of rationality, by appeal to which one logic could be seen (by all) as better than another, and they have to presume that they share this ideal, but they don’t have to be able to formulate this standard explicitly—that’s what they’re working on. And in just the same spirit, people with radically different ideas about which
policies or laws would best serve humanity can—indeed, must—presuppose some shared ideal if there is to be any point in talking it over at all.

Balkin provides an imaginary dialogue that illustrates the appeal to transcendent values in its simplest form. A marauding army massacres the people and we call them war criminals. They object, saying that their culture permits what they have done, but we can turn their point back on them.

...we can say to them: “If standards of justice and truth are internal to each culture, you can have no objection to our characterization of you as war criminals. For just as our standards can have no application to you, your standards can have no application to us. We are as correct in proclaiming your evil in our culture as you are correct in proclaiming your uprightness in yours. But your very assertion that we have misunderstood you undermines this claim. It presupposes common values of truth and justice that we are somehow obligated to recognize. And on that ground we are prepared to argue for your wickedness.” [p. 148]

This plea may fall on deaf ears, but if so, then there really are objective grounds for a verdict of irrationality: they are making a mistake that they themselves have no grounds to defend to themselves, and that we need not respect in deference.

Cultural evolution has given us the thinking tools to create our societies and all their edifices and perspectives, and Balkin sees that these thinking tools—which he calls cultural software—are inevitably both liberating and constraining, both empowering and limiting. When our brains come to be inhabited by memes that have evolved under earlier selection pressures, our ways of thinking are
restricted just as surely as our ways of talking and hearing are restricted when we learn our mother tongue. But the reflexivity that has evolved in human culture, the trick of thinking about thinking and representing our representations, makes all the restrictions temporary and revisable. As soon as we recognize that, we are ready to adopt what Balkin calls the ambivalent conception of ideology which avoids Mannheim’s paradox: “A subject constituted by cultural software is thinking about the cultural software that constitutes her. It is important to recognize that this recursion in and of itself involves no contradiction, anomaly, or logical difficulty” (pp. 127–28). Balkin insists, “Ideological critique does not stand above other forms of knowledge creation or acquisition. It is not a master form of knowing” (p. 134). This book is intended to be an instance of just such an ecumenical effort, relying on the respect for truth and the tools of truth-finding to provide a shared pool of knowledge from which we can work together toward mutually comprehended and accepted visions of what is good and what is just. The idea is not to bulldoze people with science, but to get them to see that things they already know, or could know, have implications for how they should want to respond to the issues under discussion.
restricted just as surely as our ways of talking and hearing are restricted when we learn our mother tongue. But the **reflexivity** that has evolved in human culture, the trick of **thinking** about **thinking** and **representing** our **representations**, makes all the restrictions temporary and revisable. As soon as we recognize that, we are ready to adopt what Balkin calls the ambivalent conception of ideology which avoids Mannheim’s paradox: “A subject constituted by cultural software is thinking about the cultural software that constitutes her. It is important to recognize that this recursion in and of itself involves no contradiction, anomaly, or logical difficulty” (pp. 127–28). Balkin insists, “Ideological critique does not stand above other forms of knowledge creation or acquisition. It is not a master form of knowing” (p. 134). This book is intended to be an instance of just such an ecumenical effort, relying on the respect for truth and the tools of truth-finding to provide a shared pool of knowledge from which we can work **together** toward mutually comprehended and accepted visions of what is good and what is just. The idea is not to bulldoze people with science, but to get them to see that things they already know, or could know, have implications for how they should want to respond to the issues under discussion.
A Key Part of Meta-Logic We All Share

Contradictions imply falsity.
Avoid contradictions!
And so …
And so …

- The many creeds corresponding to the many “main” religions are pairwise contradictory — a brute fact we can see when we step “above” any particular religion (including our own, if we have one).
And so …

- The many creeds corresponding to the many “main” religions are pairwise contradictory — a brute fact we can see when we step “above” any particular religion (including our own, if we have one).

- Therefore, …?
  - They can’t all be true.
  - No two can be true.
  - None are true.
  - None are likely to be true.
  - No one can be true.
  - Each is unlikely to be true.
  - It’s unlikely that any are true.
And so …

- The many interpretations corresponding to the many “main” interpretations of quantum mechanics are pairwise contradictory—a brute fact we can see when we step “above” any particular interpretations (including our own, if we have one).

- Therefore, … ?
  - They can’t all be true.
  - No two can be true.
  - None are true.
  - None are likely to be true.
  - No one can be true.
  - Each is unlikely to be true.
  - It’s unlikely that any are true.
More Sophisticated Direction?
More Sophisticated Direction?

• The mark of the vicinity of truth is a small number of contending frameworks among smart, learned people; and the mark of the vicinity of falsity is a large number of contending frameworks among people …
More Sophisticated Direction?

• The mark of the vicinity of truth is a small number of contending frameworks among smart, learned people; and the mark of the vicinity of falsity is a large number of contending frameworks among people …

• But how do you actually count the frameworks, in science and religion?
AI as (Profound) Art

Selmer Bringsjord

November 4, 2015

Presentation, December 3 2015

As computer scientists well know, the proposition that computer programming is art, and the corollary that computer programmers are artists, is one that has been widely advanced and affirmed. Confirmation of this state-of-affairs can be easily found; Exhibit A is Knuth’s monumental *The Art of Computer Programming*. But what does ‘art’ and ‘artist’ here mean? That the programs in question are elegant, efficient, bug-free, and effective; and that those who create such artifacts are praiseworthy for doing so. Exactly parallel things are often said about engineers and the best of the artifacts that they produce. Most assuredly, what is *not* meant is that programmers and engineers should strive to produce *profound* art. Profound art creates deep emotional experiences in those who perceive it, and the job of the true artist, even the great artist, is to create that which causes such experiences, repeatedly; mere utilitarian considerations are cast aside. On this metric, computer programming, and engineering in general, fall short — usually. I say ‘usually’ because a lot of AI is profound art, and is intended to be so by the artists in question. This is especially true for AI robotics. I explain that those who create such artifacts as Deep Blue, Watson, Siri, and — especially — humanoid robots (like Aldebaran’s Pepper) are *bona fide* artists striving to produce profound art.

**Knuthian Shortcut**

Reading *The Art of Computer Programming* takes a long, long time. There is a shortcut route to seeing that programming, for Knuth, is an art, but not profound art: read a short essay in CACM by Knuth, from 1974.
slutten