On *Breaking the Spell* of Irrationality; A Better Version of Pascal's Wager

Selmer Bringsjord
(with Atriya Sen & Naveen Sundar G)

Are Humans Rational?
11/21/19
RPI

Some Logistics

Some Logistics

- Recall schedule: Next *three* classes on “Steeples of Rationalistic Genius” — from Gödel.

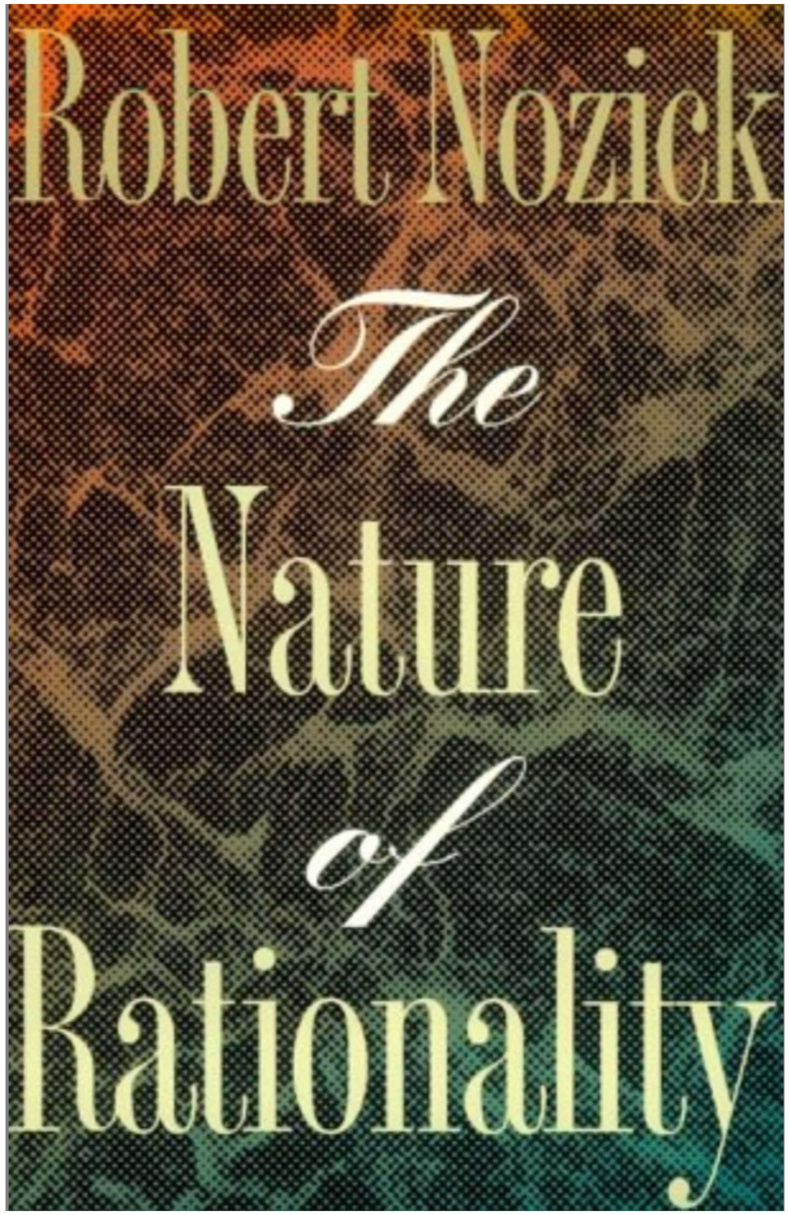
Some Logistics

- Recall schedule: Next *three* classes on "Steeples of Rationalistic Genius" — from Gödel.
- Papers due 11/25 by 5pm. (If format violated, returned without grade.)

Some Logistics

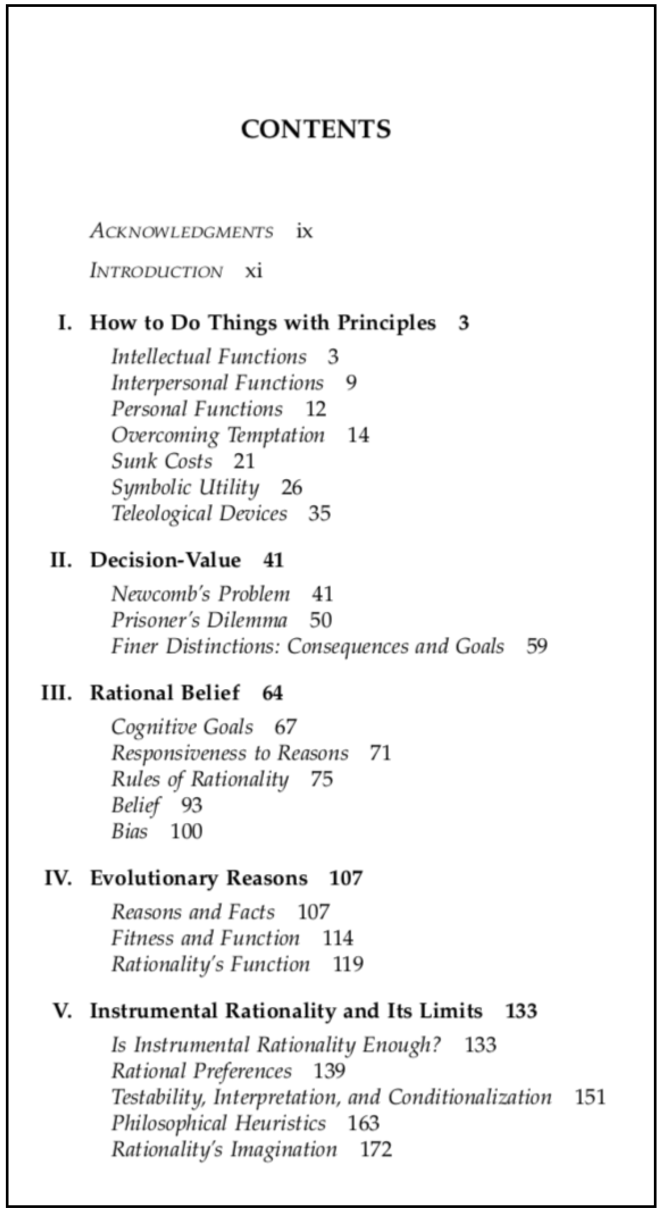
- Recall schedule: Next *three* classes on "Steeples of Rationalistic Genius" — from Gödel.
- Papers due 11/25 by 5pm. (If format violated, returned without grade.)
- Last mtg is Test #3.
  - Must understand our Gödelian coverage!
  - You can plan now to need to take *your* stand on *R-H*, or some aspect(s) thereof, in one of your essays. And you will need to anticipate and rebut at least one powerful objection to your stand/argument.

For those writing on Newcomb's Problem: Pollock & …



Robert Nozick

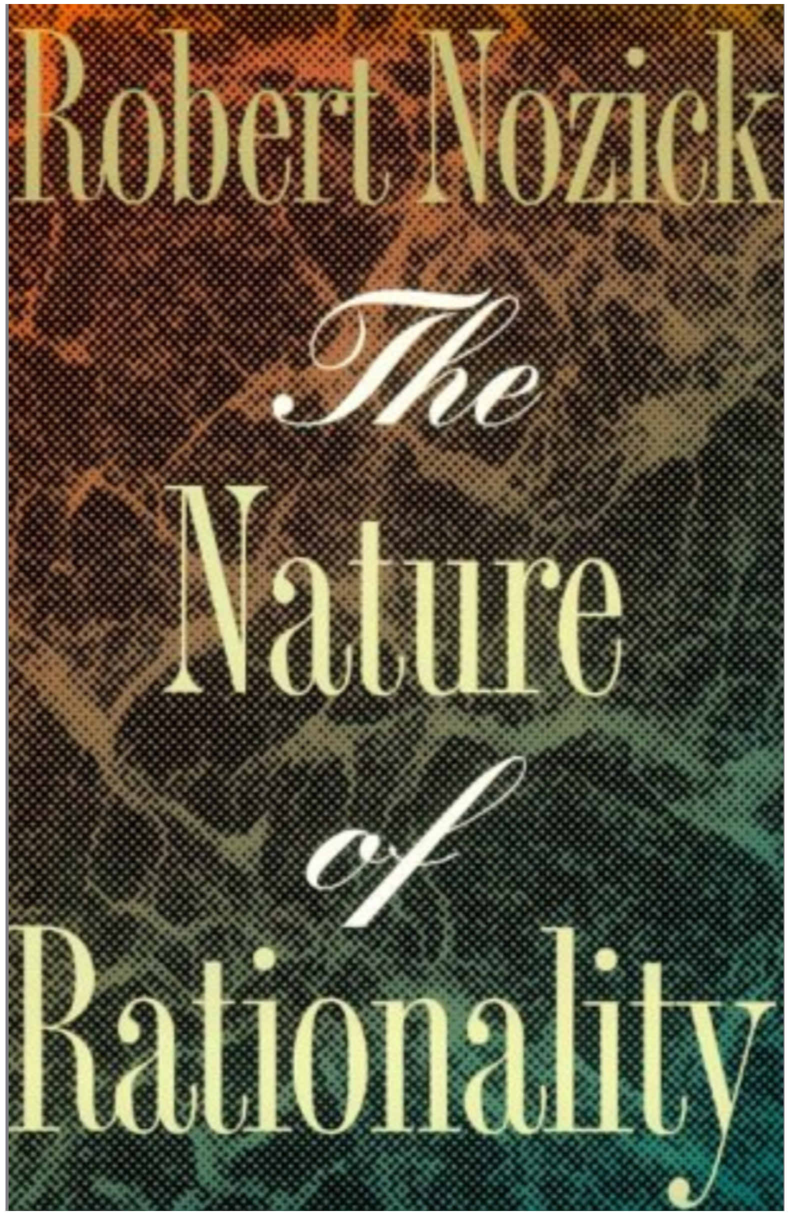
The Nature of Rationality

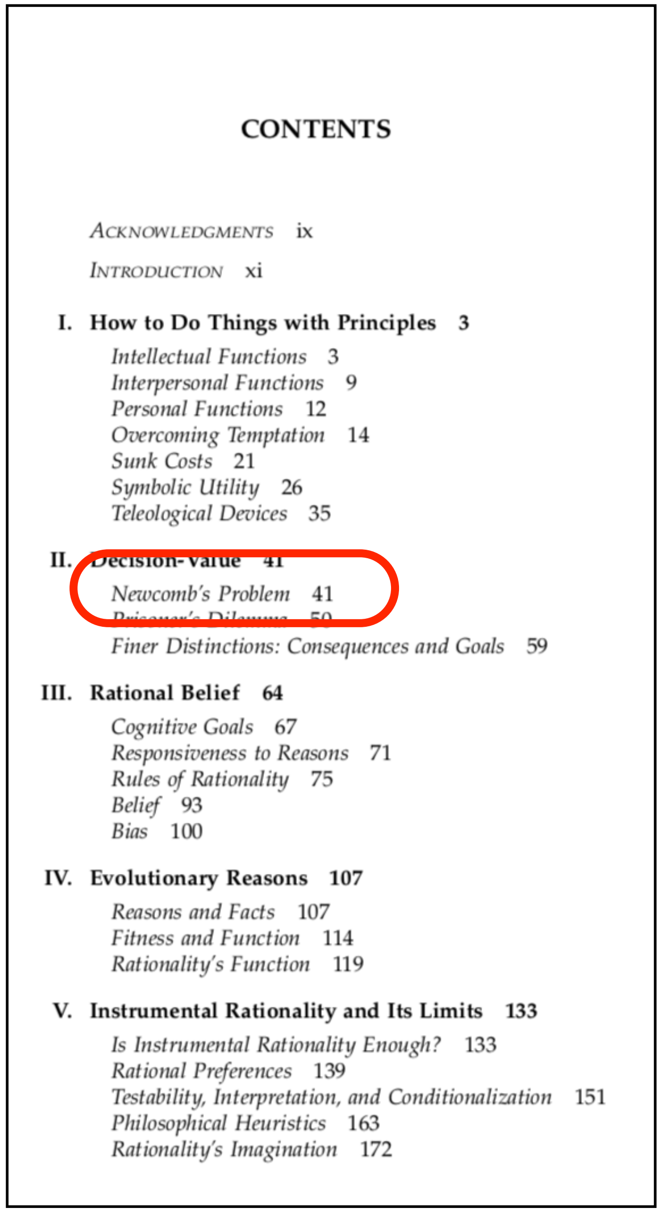


CONTENTS

http://www.univpgri-palembang.ac.id/perpus-fkip/Perpustakaan/American%20Phylosophy/Nozick%20R.%20The%20Nature%20of%20Rationality.pdf

For those writing on Newcomb's Problem: Pollock & …





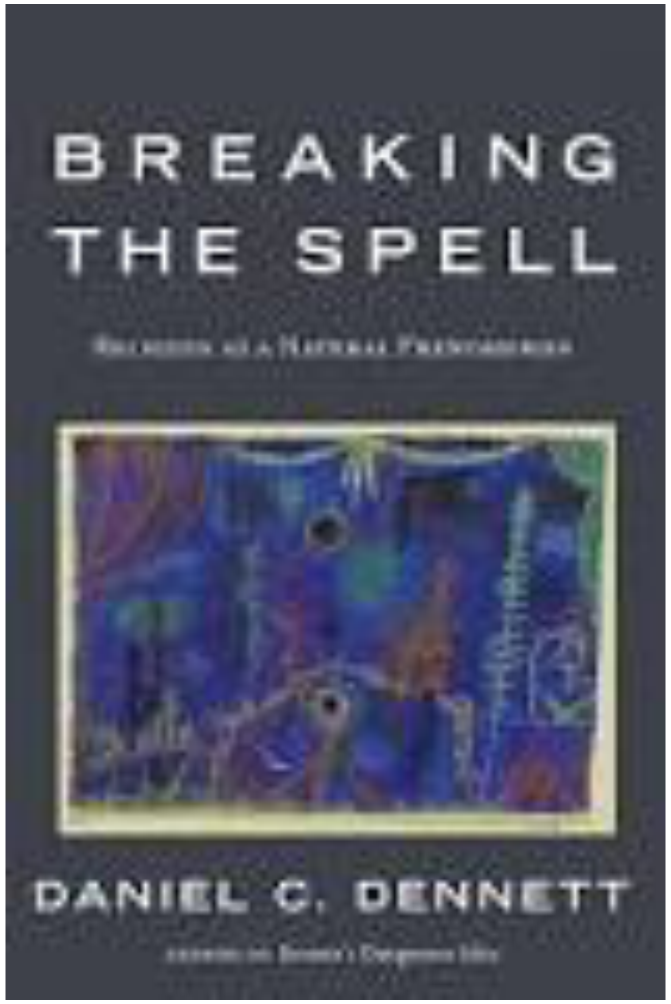
CONTENTS

http://www.univpgri-palembang.ac.id/perpus-fkip/Perpustakaan/American%20Phylosophy/Nozick%20R.%20The%20Nature%20of%20Rationality.pdf

On Religion & Rationality …

versus

The Book


BREAKING THE SPELL
Religion as a Natural Phenomenon
DANIEL C. DENNETT
Author of Darwin's Dangerous Idea

Found this on W3: http://skepdic.ru/wp-content/uploads/2013/05/Daniel_C_Dennett_Breaking_the_Spell_Religion.pdf

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:

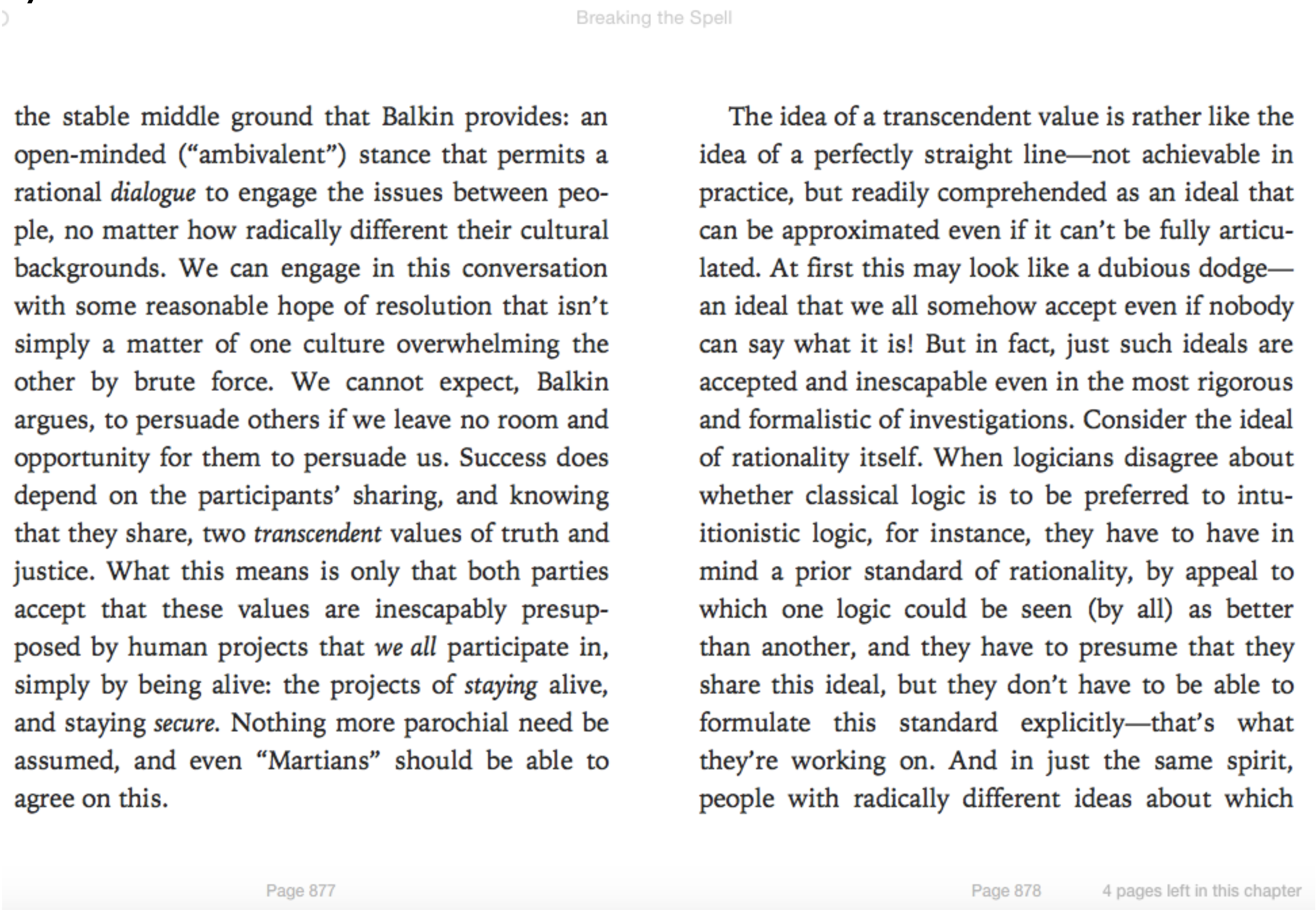
- Theists and atheists share an affirmation of, and both in fact use, a common thing: *thinking tools* (= "cultural software") that cut(s) across all human beings.

- 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 PHP & 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 PHP & 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!

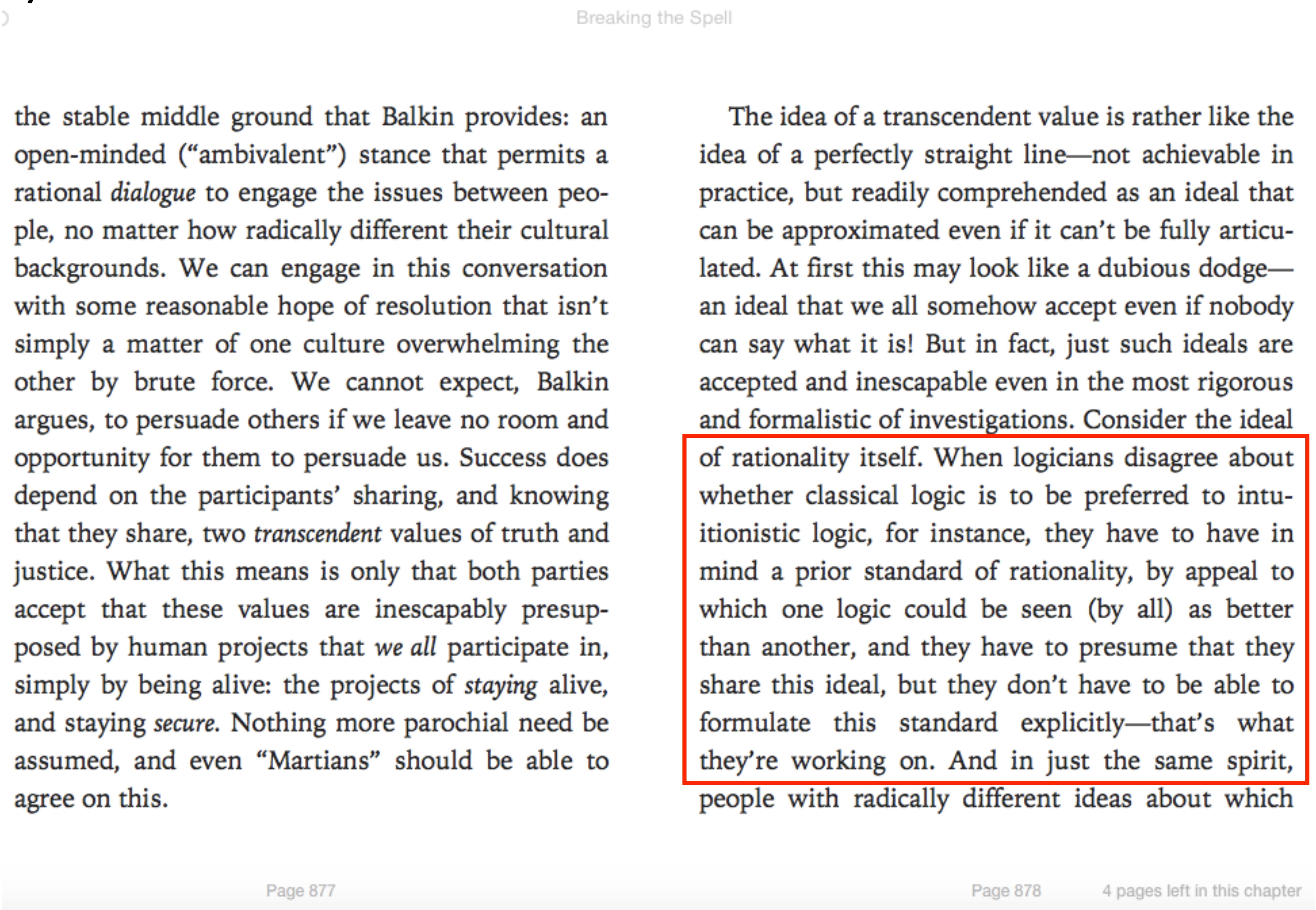
Key Text in BTS



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

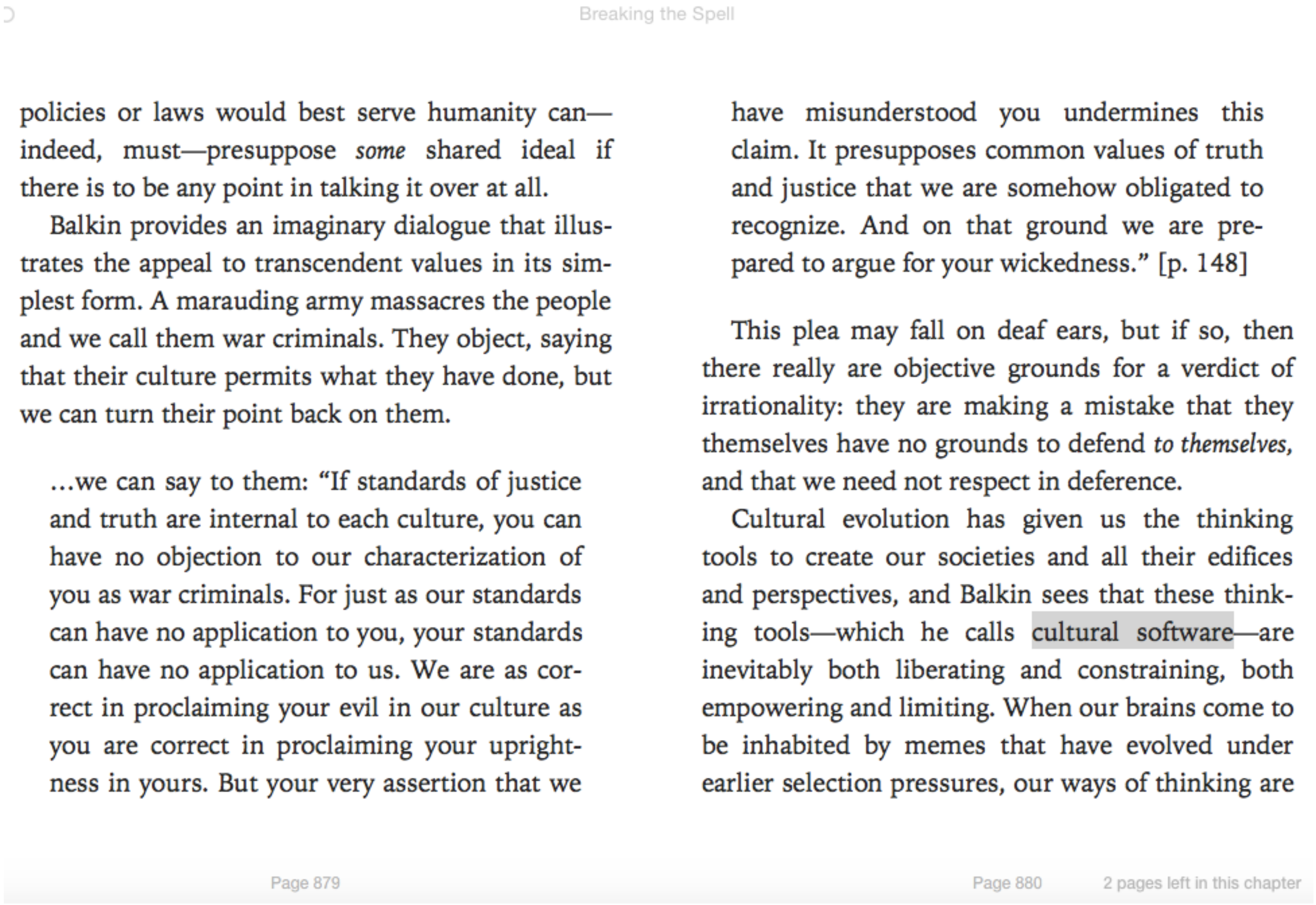
Key Text in BTS



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

Key Text in BTS



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.

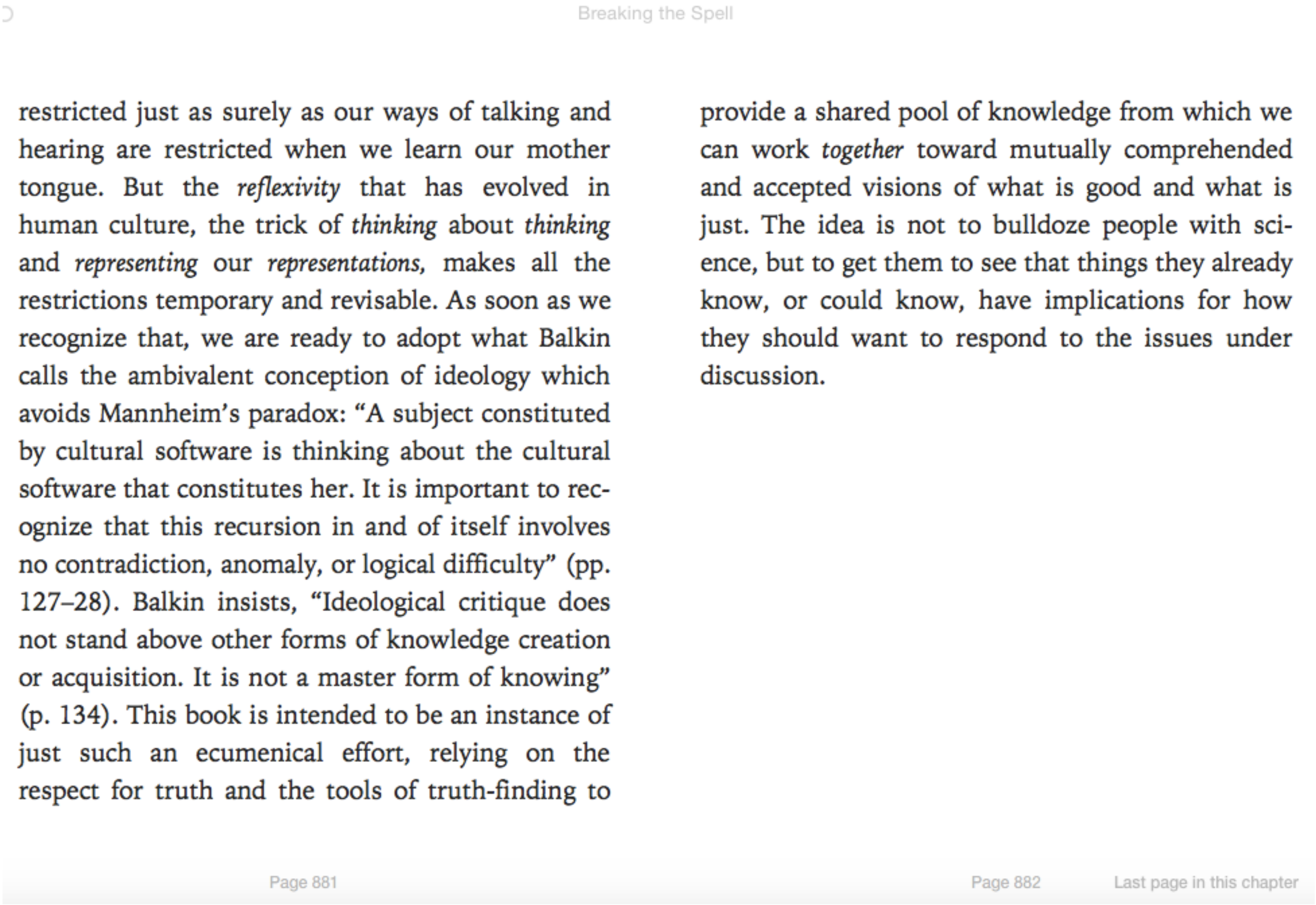
Balkin provides an imaginary dialogue that illustrates the appeal to transcendent values in its simplest form. A marauding army massacres the people and we call them war criminals. They object, saying that their culture permits what they have done, but we can turn their point back on them.

> ...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.

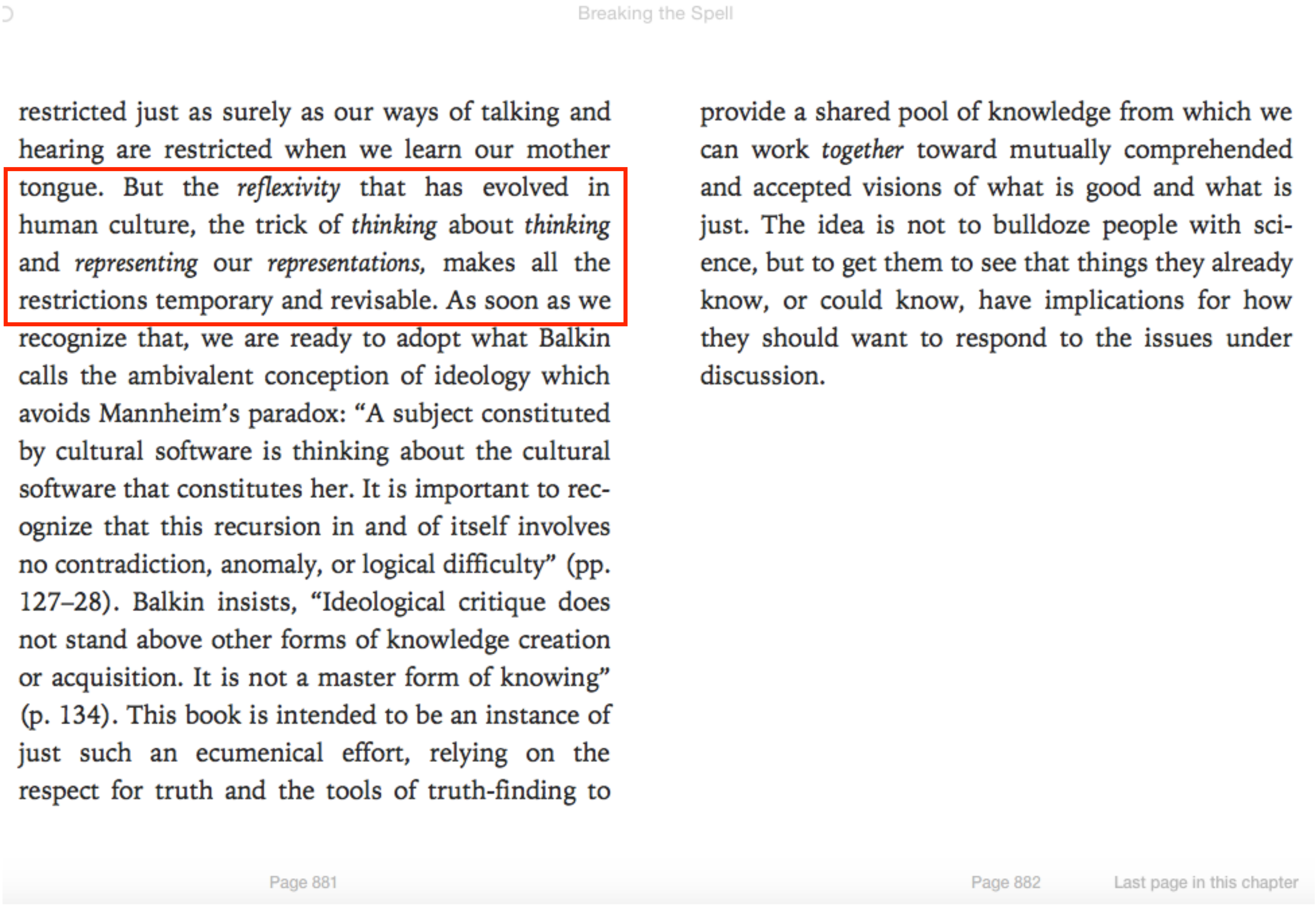
Cultural evolution has given us the thinking tools to create our societies and all their edifices and perspectives, and Balkin sees that these thinking tools—which he calls cultural software—are inevitably both liberating and constraining, both empowering and limiting. When our brains come to be inhabited by memes that have evolved under earlier selection pressures, our ways of thinking are

Key Text in BTS



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.

Key Text in BTS



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 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. Dennett
  - 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 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. Dennett
  - None are likely to be true.
  - No one can be true.
  - Each is unlikely to be true.
  - It's unlikely that any are true.

But this inference is illogical, and hence irrational.

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. Dennett
  - None are likely to be true.
  - No one can be true.
  - Each is unlikely to be true.
  - It's unlikely that any are true.

But this inference is illogical, and hence irrational.

My, *that's* ironic.

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. Dennett
  - None are likely to be true.
  - No one can be true.
  - Each is unlikely to be true.
  - It's unlikely that any are true.

But this inference is illogical, and hence irrational.

My, *that's* ironic.

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. Dennett
    - None are likely to be true.
    - No one can be true.
    - Each is unlikely to be true.
    - It’s unlikely that any are true.

But this inference is illogical, and hence irrational.

My, *that’s* ironic.

After all, consider …

After all, consider …

- 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).

After all, consider …

- 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.

After all, consider …

- 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.

After all, consider …

- 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.

Btw …

https://www.ted.com/talks/dan_dennett_let_s_teach_religion_all_religion_in_schools/transcript?language=en

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?

A *Better* Pascal's Wager

. . .

https://plato.stanford.edu/entries/pascal-wager/

Pascal's Decision Matrix (= **M**)

	G	**not-G**
Bet on G	∞	v_1
Bet on not-G	v_2	v_3

where background propositions include
'if G, then repentance secures infinite bliss etc.'.

The Optimality Principle$_2$ (OP$_2$)

(recall from coverage of Newcomb's Paradox)

When choosing between alternative actions a_1 and a_2, rationality dictates choosing that action that maximizes expected value, computed by multiplying the value of each outcome that can result from each action by the probability that it will occur, adding the results together, and selecting the action associated with the higher utility.

The Optimality Principle$_2$ (OP$_2$)

(recall from coverage of Newcomb's Paradox)

When choosing between alternative actions a_1 and a_2, rationality dictates choosing that action that maximizes expected value, computed by multiplying the value of each outcome that can result from each action by the probability that it will occur, adding the results together, and selecting the action associated with the higher utility.

(As we said before:

This principle is taught to students in every introductory economics or decision-theory class, and is at least usually a key thing to follow in the pursuit of rational behavior.)

13-Strength-Factor Continuum

Certain

Evident

Overwhelmingly Likely

Beyond Reasonable Doubt

Likely

More Likely Than Not

Counterbalanced

More Unlikely Than Not

Unlikely

Beyond Reasonable Belief

Overwhelmingly Unlikely

Evidently False

Certainly False

13-Strength-Factor Continuum

Certain

Evident

Overwhelmingly Likely

Beyond Reasonable Doubt

Likely

More Likely Than Not

Counterbalanced

More Unlikely Than Not

Unlikely

Beyond Reasonable Belief

Overwhelmingly Unlikely

Evidently False

Certainly False

13-Strength-Factor Continuum

Certain

Evident

Overwhelmingly Likely

Beyond Reasonable Doubt

Likely

More Likely Than Not

Counterbalanced

More Unlikely Than Not

Unlikely

Beyond Reasonable Belief

Overwhelmingly Unlikely

Evidently False

Certainly False

13-Strength-Factor Continuum

Epistemically Positive

Certain

Evident

Overwhelmingly Likely

Beyond Reasonable Doubt

Likely

More Likely Than Not

Counterbalanced

More Unlikely Than Not

Unlikely

Beyond Reasonable Belief

Overwhelmingly Unlikely

Evidently False

Certainly False

13-Strength-Factor Continuum

Epistemically Positive

Certain

Evident

Overwhelmingly Likely

Beyond Reasonable Doubt

Likely

More Likely Than Not

Counterbalanced

More Unlikely Than Not

Unlikely

Beyond Reasonable Belief

Overwhelmingly Unlikely

Evidently False

Certainly False

Epistemically Negative

13-Strength-Factor Continuum

Epistemically Positive

Certain

Evident

Overwhelmingly Likely

Beyond Reasonable Doubt

Likely

More Likely Than Not

Counterbalanced

More Unlikely Than Not

Unlikely

Beyond Reasonable Belief

Overwhelmingly Unlikely

Evidently False

Certainly False

Epistemically Negative

13-Strength-Factor Continuum

Epistemically Positive

(12) Certain

(11) Evident

(10) Overwhelmingly Likely

(9) Beyond Reasonable Doubt

(8) Likely

(7) More Likely Than Not

(6) Counterbalanced

(5) More Unlikely Than Not

(4) Unlikely

(3) Beyond Reasonable Belief

(2) Overwhelmingly Unlikely

(1) Evidently False

(0) Certainly False

Epistemically Negative

An Optimality Principle (OP_2*)

(based on 13-valued scheme used in solving the Lottery Paradox, St Petersburg Paradox, …)

When choosing between alternative actions a_1 and a_2, rationality dictates choosing that action that maximizes expected value, computed by multiplying the value of each outcome that can result from each action by the *likelihood* (0 to 13) that it will occur, adding the results together, and selecting the action associated with the higher utility.

A rational person must bet that God exists.
–B. Pascal

A rational person must bet that God exists. –B. Pascal

Proof: We employ that any natural (or, for that matter, real) number *n* multiplied by/added to an infinite utility value yields an infinite utility value (unless $n = 0$). We observe that the likelihood God exists is at minimum *evidently false* (1).$^{++}$ But then the expected utility value of betting on G is infinite, whereas the expected utility value of betting that God doesn't exist is finite. (Why, exactly?) Hence, by OP_2* a rational agent will bet on G (i.e. bet that God exists). **QED**

A rational person must bet that God exists.
–B. Pascal

Proof: We employ that any natural (or, for that matter, real) number *n* multiplied by/added to an infinite utility value yields an infinite utility value (unless $n = 0$). We observe that the likelihood God exists is at minimum *evidently false* (1).++ But then the expected utility value of betting on G is infinite, whereas the expected utility value of betting that God doesn't exist is finite. (Why, exactly?) Hence, by OP_2* a rational agent will bet on G (i.e. bet that God exists). **QED**

++Oxford's Richard Swinburne has a large body of work designed to show that *prob*(G) is at minimum greater the .5; i.e. — in my likelihood framework, at least *more likely than not*.

slutten