

Dismantling the Guillotine

PHIL ROBERTS, JR.
www.rationology.net

Abstract: Aside from the mind body problem, there is perhaps no more infamous philosophical perplexity than ‘Hume’s Guillotine’, i.e., the purported logical gap between ‘is’ (factual/ descriptive) premises and ‘ought’ (normative/ prescriptive) conclusions. Based on the premise that our prescriptive ‘ought’s are actually referencing an implicit theory of rationality along with my own proposed procedure for deciphering some of its parameters, I offer a derivation of a moral prescriptive ‘ought’ (‘Love thy neighbor’) from a descriptive ‘is’ (an implicit theory of rationality that is demonstrably “true”) on the grounds that the concept/ attribute of rationality is the fount from which all normativity flows. I also offer an explanation for why we, as quasi rational naturally selected organisms experience and often respond to the tug of this moral maxim (‘Love thy neighbor’) including those occasions when doing so falls well outside the predicted parameters of inclusive fitness theory (Mother Teresa, self-endangering Greenpeacers, self-incinerating Buddhist monks, etc.).

1. Background

At our 25th annual meeting (Washington, D.C.), I presented a theory of emotional instability (Roberts, 2005) based on a simple premise, i.e., the premise that ‘feelings of worthlessness’¹ are not so much an adaptation as a part of the price we humans have had to pay for having become a little too rational/ objective for our own good -- a maladaptive byproduct of the evolution of rationality that Mother Nature “tolerates” as a necessary premium for having a rational species to do her bidding.²

The gist of the argument here was that, as humans became smarter and smarter about how the world is put together – presumably transpiring over millennia of linguistic and cultural evolution -- they gradually became smarter and smarter about values. That is to say, contrary to Hume’s famous dictum (1739) that reason is strictly a matter of truth and falsehood, they gradually became increasingly adept at distinguishing between values that are justifiable (e.g., supported by evidence)³ and those that are simply a matter of happenstance. Eventually, so this story goes, this culminated in an increased volatility in the most crucial value of all – one that in all likelihood lies at the very heart of the will to survive – an increased volatility in self-value along with a host of maladaptive effects (anxiety, depression, addiction, suicide, guilt, etc.). Stated in the form of an outline of a theory, I began this outline as follows:

Objective: To answer the question: ‘Why are the members of one particular species of naturally selected organism expending significant amounts of effort and energy on the biologically bizarre non-physical objective of maximizing self-worth?’⁴

Explanation: Being the blind arational process that she is, Mother Nature instills in all her creatures a sense of their own importance (or of the importance of their needs) that is rationally inordinate. And, as a species reaches a certain stage in its rational/ cultural/ memetic development, its members increasingly come to question this inordinacy, and increasingly come to require reasons (justification) for maintaining it (needs for love, purpose, meaning, moral integrity, autonomy, justice, etc.).

The implications of this conjecture are extensive, including implications with respect to emotional disorder, free will,⁵ rationality, ethics and incompleteness:

Incompleteness: When viewed from the context of our underlying premise, ‘feelings of worthlessness’ can be construed as providing an empirical vindication of the Lucas/ Penrose assertion that Gödel’s theorem (1931) implies that “minds are different from machines” (Lucas, 1961). The empirical vindication of this much contested assertion is based on the following argument:

1. ‘Feelings of worthlessness’ constitute evidence that humans are beginning to question the value of their existence and therefore are beginning to question the value of achieving the fixed objective of Mother Nature’s most basic program (survival).⁶
2. The same capacity for “standing outside the system” (Lucas) that enables us to “see” that the Gödel sentence is “true” is what enables us to stand outside of nature’s most basic program and question (in the guise of ‘feelings of worthlessness’) whether it is one worth completing.
3. Rationality cannot be constrained (captured in its entirety) within a formal system, in this case, one that has been “designed” to achieve a fixed objective, not even by Mother Nature herself.

Although this argument assumes at least some prior familiarity with the relevant literature (Penrose, Hofstadter, Lucas, etc.), the important point here is that, in a world where my premise is correct and I have not misread its implications, one would have reason to suspect that all concrete instantiations of rationality are likely to be incomplete and therefore that no individual, culture, belief, theory, value, objective, etc. is likely to be rational in any but a quantitatively relative sense of the term (e.g., X is relatively more or less rational than Y, the norm, etc.).⁷

In what follows I wish to explore another one of the implications of my conjecture, what I perceive to be its implication in the realm of ethics:

Ethics: Since, according to the above explanation, ‘being more rational’ entails ‘being more valuatively objective/ impartial’, the moral maxim, ‘Love (i.e., intrinsically value) your neighbor as you love yourself’ could be construed as an imperative of an implicit theory of rationality in which, among other things, ‘being rational’ entails ‘being valuatively objective/ impartial’. This would also mean that, to the extent this implicit theory turns out to be “true”, the author of Genesis actually got it right in referring to our awareness of right and wrong as a form of knowledge (moral realism).

2. Being Rational

If one assumes, as I have conjectured, that feelings of worthlessness are a maladaptive byproduct of the evolution of rationality in the sense of resulting from a more objective understanding of how the world is put together, it seems to follow that the "standard picture" of rationality (Stein, 1996) is mistaken. 'Being rational' is not so much a matter of slavishly conforming to established rules of inference (a process) as a matter of 'being able to "see" what is going on' as a result of reasoning that has already transpired, whether one's own, or culturally acquired – a state of mind facilitated by the product of a process.⁸ This would also mean that, rather than an assessment of one's reasoning, our common sense rationality ascriptions might better be construed as appraisals of a mental map of sorts, one apparently comprised of both beliefs and values, in which the cognitive component of this "seeing" correlates with the extent to which the map is comprised of beliefs that accurately and coherently represent reality.⁹ Among many other things, this is taken to include beliefs about how to acquire beliefs that accurately and coherently represent reality, reflected in how well one reasons.

Finally, since in this view no individual is likely to be rational in any but a relative sense of the term (incompleteness), this would also mean that when we refer to an individual as "rational" or "irrational" that we are probably just expressing a rough appraisal of how this individual's mental map compares to the norm in facilitating this ability to "see". We mistakenly employ bivalent terminology -- terminology that would be better suited for logical assertions -- to express what are actually relational judgments or evaluations.¹⁰

Summarizing:

'Being rational' entails 'being able to "see" what is going on' or, less metaphorically, 'being objective', not only cognitively, but valuatively as well (e.g., impartiality) and in which no individual, culture, belief, theory, value, objective, etc. is likely to be rational in any but a quantitatively relative sense of the term (e.g., X is relatively more or less rational than Y, the norm, etc.).

3. Virtues of the Implicit Theory

In Roberts, 2006, presented at our 26th annual meeting (Bloomington), I detailed some of the explanatory advantages of the above theory/ definition (expressed in the idiom of ‘being rational’) which I will refer to as the “implicit theory”. For example, the implicit theory...

- a. can explain what morality “is” (p. 2),
- b. can unify the sexual and altruistic aspects of morality (pp. 2, 3),
- c. exhibits logical symmetry between the valuative and cognitive domains (p. 3),
- d. has much in common with a longstanding paradigm for rationality (pp. 3, 4),
- e. can resolve the paradox of rational irrationality (pp. 4,5),
- f. can ameliorate theoretical vs. practical rationality conflict (p. 5),
- g. can resolve the “rationality debate” in a straightforward manner (p. 5),
- h. is immune to the paradox of the Prisoner’s Dilemma (p. 6),
- i. is compatible with our common sense understanding of jury selection (p. 6),
- j. can afford an avenue for the justification of a moral ‘ought’ (pp. 6,7),
- k. can resolve the Gettier problem (forthcoming),
- l. can endorse or provide a metajustification for explanationism (forthcoming).

To illustrate with respect the valuative objectivity feature (e.g., item h.), in a standard Prisoner’s Dilemma in which each prisoner will be individually better off by defecting but both will be better off by co-operating, typical game theoretic notions of rationality prescribe a course of action that is less than optimal (defection) which, for many theoreticians, has seemed counterintuitive. In contrast, in the world of the implicit theory, the more rational the actors become the more they will ‘be able to “see” what is going on’, not only with respect to the optimal strategy (co-operation), but also with respect to an increased concern for each other’s well being (an increase in valuative objectivity/ impartiality) in which the optimal strategy will ensue, even if only as an unintended consequence.

To illustrate with respect to the incompleteness feature (e.g., item g.), the avalanche of experimental evidence that humans systematically violate established rules of inference (the heuristics and biases research initiated by Tversky and Kahneman, 1974) could not be construed as evidence that ordinary humans are irrational, as scores of experts on the subject have seemed to maintain (‘Open Peer Commentary’ in Cohen, 1981). Nor would it be necessary to navigate the intricacies of a wide versus a narrow reflective equilibrium or to introduce a competence/ performance distinction to establish why this might be so (Cohen, 1981, Stein 1996, etc.).¹¹ This is because, while all ascriptions are relational of necessity (incompleteness), they are relational relative to the norm simply as a matter of custom and convenience. As such, there is nothing to preclude us from inferring from the experimental evidence that expert opinion might be relatively more rational than the norm (in terms of ‘being able to “see” what is going on’) where such matters are concerned,¹² while at the same time preserving our unshakeable conviction regarding the sanctity of the rationality of ordinary humans.

Any notion that humans on the whole are irrational in significant, systematic ways would seem, well, irrational to most of us (Renee Elio).

4. Dismantling the Guillotine

Aside from the mind/ body problem, there is perhaps no more infamous philosophical perplexity than ‘Hume’s Guillotine’ (Black, 1964), i.e., the purported logical gap between ‘is’ (factual/ descriptive) premises and ‘ought’ (normative/ prescriptive) conclusions (Hume, 1739):

In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the copulations of propositions, 'is', and 'is not', I meet with no proposition that is not connected with an 'ought', or an 'ought not'. This change is imperceptible; but is, however, of the last consequence...¹³

Given the implications I have presented in sections 1 and 2, it should come as no surprise that I believe there is considerably more to our moral admonitions than our spoken words – that they are underwritten or endorsed by an implicit (e.g., subconscious) “theory” of rationality. And, if I am right about this, then it might indeed be possible to derive a moral ‘ought’ -- the kind I assume Hume was specifically targeting given his disdain for “vulgar systems of morality” -- from a descriptive ‘is’ by...

1. deciphering at least some of the parameters of our “implicit theory” (Sections 1 and 2),
2. establishing that this implicit theory is “true” by demonstrating that it can maximize explanatory coherence better than competing theories of rationality (Section 3) and...
3. deriving a prescriptive ‘ought’ based on the values described by this implicit theory (the descriptive ‘is’), not on the grounds that rationality is a valued end, rendering the ‘ought’ instrumental, but rather on the grounds that it is only within the realm of rational thought and discourse underwritten by this shared implicit theory that argumentative, persuasive, prescriptive ‘ought’s are possible (a categorical prescriptive ‘ought’).

Fleshing out some of the details, given the specific theory of ‘being rational’ I have summarized at the conclusion of Section 2 we would then have:

STEP THREE:

Given that it is demonstrably “true” that ‘being rational’ entails ‘being objective’, not only cognitively, but valuatively as well (impartiality), it is demonstrably true that, ideally,¹⁴ one ‘ought’ to ‘Love (intrinsically value) one’s neighbor as one’s self’ on the grounds that the concept/ attribute of rationality is the fount from which all normativity flows.

Lest there be any remaining confusion here, my tactic has been to justify a widely accepted moral maxim by equating morality with rationality -- by bringing the mountain to Mohammed, so to speak.

There is a reason to do something just in case it is rational to do it. That is why, in some contexts, there can be a need to justify morality or law or even prudence, but no similar need for a justification of rationality (Dreier, 2001)

5. Why We Care

A successful theory of morality should make it clear how the wrongness feature should be practically relevant to me – why it should move me (Scott James).

Envision a scenario in which, in the heat of battle, a soldier succumbs to the all but irresistible urgings of the lower emotion of fear and, as a result, fails to do his utmost to prevent his fellow soldiers from having to pay the ultimate price. And imagine that in this scenario, which I assume is not all that atypical for myriads of individuals placed in a similar predicament, the soldier physically survives only to live out his remaining days in a drunken stupor as a result of his inability to “live with himself”.

Rather than straining credulity to the breaking point by trying to argue that, aside from the decision to defect, this type of scenario plays out in a manner that is actually adaptive,¹⁵ the explanation I offered in my 2005 outline was considerably more nuanced:

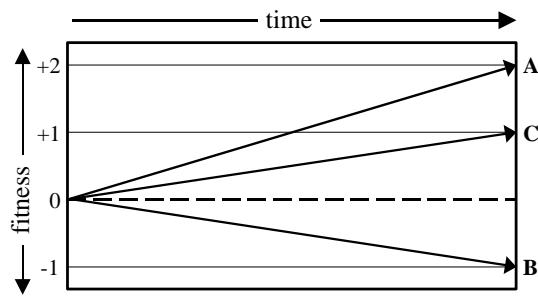
Moral Sentiment: If the emergence of an awareness of right and wrong is simply an emerging awareness of the nature of rationality itself, then it would also provide a rationale for the persuasive force of moral argument, in that it could be construed as exploiting our need to view ourselves as rational as a determinant in assessing self-worth. Although, as a part of the basis for assessing self-worth, none of us can actually measure up to the standard of loving others as we love ourselves (valuative objectivity/impartiality), in this view we nonetheless come to experience feelings of worthlessness (guilt) along with a corresponding reduction in the will to survive (depression, addiction, suicide, etc.) when we deviate from the standard to an unacceptable degree. In other words, a capacity for guilt (having a conscience) is a part of the price we humans have had to pay for having become a little too rational/ objective for our own good – a maladaptive (from the standpoint of the individual or “gene”) manifestation of our need to justify our existence which Mother Nature “tolerates” as a necessary premium for having a rational species to do her bidding.¹⁶

We struggle to be honest and courteous and responsible and brave in circumstances where it is difficult. Even if apes are sometimes courteous, responsible, and brave, it is not because they think they should be. Even as primitive a phenomenon as a teenager's efforts to be "cool" is a manifestation of the human tendency to live a life guided by ideals rather than merely driven by impulses and desires. We also suffer deeply from our self-evaluations and act in sick and evil ways as a result. This is part of what I had in mind earlier when I said that human beings seem psychologically damaged in a way that suggests a break with nature (Christine Korsgaard).

Endnotes

1. In contrast with those who equate naturalism with materialism, I have never thought there was anything even remotely “scientific” about the assumption that the mind/ body problem has somehow been miraculously resolved as a result of the admittedly amazing progress in the material sciences. As such, it should come as no surprise that I am assuming that ‘feelings of worthlessness’ are a non-physical feature of nature – one produced by millions of years of neurological evolution operating in tandem with millennia of linguistic and cultural evolution – that are probably exclusively present in enculturated human minds. The wild boy of Aveyron and, prior to her famous Aha!, Helen Keller, no doubt experienced lots of feelings (pain, fear, aggression, frustration, despair, etc.), but I suspect that feelings of worthlessness were probably not among them.
2. In somewhat more literal terms, the maladaptiveness of an increased volatility in the will to survive (feelings of worthlessness, suicide, etc.) is outweighed by the adaptiveness of a massive increase in the ability to survive (agriculture, technology, global relief, etc.):

In its most simplistic formulation, accounting for the sustained presence of a mechanism presumed to produce maladaptive values (deviations from the predicted profile of “ruthless selfishness” – Dawkins, 1976) is simply a matter of assuming that developments in our cognitive profile have enhanced our ability to survive to such a degree that it more than compensates for the increased volatility in the resolve to do so (Figure 8) (Roberts, 2001).



A = benefits of cognitive developments
B = costs of ensuing evaluative developments
C = net gain in fitness

Figure 8: Sustaining the mechanism

3. Hume, himself, offers an example of “objective evidence” concerning one’s worth in his *Treatise of Human Nature*, Book II, Part I, Section XI:

Tis now time to turn our view from the general consideration of sympathy, to its influence on pride and humility, when these passions arise from praise and blame, from reputation and infamy..... Now nothing is more natural than for us to embrace the opinions of others in this particular; both from sympathy, which renders all their sentiments intimately present to us; and from reasoning, which makes us regard their judgment, as a kind of argument for what they affirm [my underline]. These two principles of authority and sympathy influence almost all our opinions; but must have a peculiar influence, when we judge of our own worth and character....

4. This should not be conflated with the assertion that humans are expending significant amounts of effort and energy on the conscious pursuit of self-worth, but merely that this need is ubiquitous:

There is no value-judgment more important to man -- no factor more decisive in his psychological development and motivation -- than the estimate he passes on himself. This estimate is ordinarily experienced by him, not in the form of a conscious, verbalized judgment, but in the form of a feeling, a feeling that can be hard to isolate and identify because he experiences it constantly: it is part of every other feeling, it is involved in his every emotional response. ... it is the single most significant key to his behavior (Nathaniel Branden).

5. As continued in the outline (Roberts, 2005):

Indeterminism: "Free will" and the incessant activity presumed to emanate from it is simply the insatiable appetite members of our species have for self-significating/ self-worth enhancing experience (juxtaposed with the need to avoid the pain of feelings of worthlessness, irrelevance, etc.) which, in turn, is simply nature's way of attempting to counter the objectifying influence of our rational faculties. As such, although "free will" itself (the self-worth complex) is constrained within parameters determined by natural selection (the maximizing of self-worth), its presence in us, manifested in the need to expend significant amounts of effort and energy on maintaining emotional well-being (keeping up with the Joneses, climbing Mt. Everest, presenting papers at conferences, etc.) would, according to this perspective, be construed as evidence that members of nature's most rational species have become too evaluatively objective (requiring remedial measures) and, as such, less evaluatively/ conatively determined by natural selection than members of less rational more emotionally stable species. In this view, indeterminism is manifested, not in the ability to change one's mind about what to have for breakfast, but rather in a species whose members are becoming less and less concerned with staying alive (e.g., daily suicide bombings in the Middle East) and more and more concerned with reasons (justification) for staying alive (needs for love, purpose, meaning, etc.).

6. Technically, it's a matter of surviving long enough to successfully reproduce and with "special" and "limited" exceptions (Dawkins, 1976) in deference to inclusive fitness theory (Hamilton, 1964).
7. Amusingly, this would also seem to suggest that I am proposing a theory of rationality that implicitly predicts its own eventual demise in that, it too, can only be construed as relatively rational.
8. If called upon for a literal definition for the term "rationality", the devil is most definitely in the details. As a first approximation, I would opt for 'the psychical (non-physical) product of ampliative inference', most of which is acquired from one's culture, and with the term "ampliative" unpacked to refer to inferences that serve to maximize the objectivity, understanding, impartiality, wisdom, etc. of the agent or agents so endowed. However, since we usually employ the term in an evaluative context, it gets messy, particularly when you take into consideration that these evaluations or assessments are relational in nature, as explained in the last paragraph of section 2. In this sense, are you ready for this, the term refers to 'the extent to which one's psychical product of ampliative inference serves to maximize one's objectivity (one's mental ocularity) relative to the norm'.

The concept of rationality, one might say, is incorrigibly elusive... I believe it is fair to say that in philosophical discussions of rationality, there is a sense in which we do not "know what we are talking about" and can never do so, if what is demanded is a concise definition (Max Black).

9. At last year's annual meeting (Bloomington), I suggested that we visualize this psychical product in terms of a follow-the-dots diagram in which "seeing" is facilitated by the extent to which the lines have been connected correctly and completely and in which cohering lines count for more than dispersed lines in terms of 'being able to "see"' what the diagram represents. I also proposed that the incompleteness feature be represented in terms of a diagram with no definite boundaries and/ or no limitation on the amount of possible detail and that, given the heavy reliance on the metaphor of vision, and its intended purpose of addressing deficiencies in the "standard picture" of rationality (Stein, 1996), that we refer to this visual representation as the "picture-picture" (Roberts, 2010).

- I do not take departures from the ordinary use of words lightly. However, I also do not construe ordinary usage as sacrosanct, particularly if a plausible rationale can be offered for why it might be mistaken such as that relational terminology is simply too cumbersome to be employed for the purposes of every day conversation.

Our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our arm-chairs of an afternoon (J. L. Austin).

- As with my brief excursion into incompleteness in Section 1, for the sake of brevity my discussion of the Prisoner's Dilemma and the heuristics and biases program in Section 3 also assumes at least some prior familiarity with the relevant literature.
- This can be a two edged sword however in that, given our common sense understanding that rationality and logic are by no means synonymous, and our common sense understanding that, come hell or high water, ordinary human beings are the standard for 'being rational', some might argue that expert opinion has been relatively less rational than the norm with regard to its longstanding love affair with the "standard picture", i.e., the view that 'being rational' is simply a matter of slavishly conforming to established rules of inference a la Stein, 1996. Indeed, some might even go so far as to suggest that, as is so often the case, perhaps the lover in question may have been blinded by lust, on this occasion, the lust to reduce mind to matter via the reduction of rationality to logic, rules, processes, principles, procedures, etc. (mechanistic materialism, computationalism, homuncular functionalism, massive modularity, etc.) via the "standard picture".
- This appears in Hume's, *A Treatise of Human Nature*, 1739, Book III, Part I, Section I:

In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the copulations of propositions, 'is', and 'is not', I meet with no proposition that is not connected with an 'ought', or an 'ought not'. This change is imperceptible; but is, however, of the last consequence. For as this 'ought', or 'ought not', expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given; for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. But as authors do not commonly use this precaution, I shall presume to recommend it to the readers; and am persuaded, that this small attention wou'd subvert all the vulgar systems of morality, and let us see, that the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceiv'd by reason.

- The stipulation "ideally" in the derivation is necessitated by the supposition that all concrete instantiations of rationality are likely to be incomplete.
- I have deliberately envisioned a scenario that should prove problematic for adaptionist accounts of morality:

Suppose there was a realm of action of such recurrent importance that nature did not want practical success to depend on the frail caprice of ordinary human practical intelligence. That realm might, for example, pertain to certain forms of cooperative behavior toward one's fellows. The benefits that may come from co-operation -- enhanced reputation, for example -- are typically long-term values and merely to be aware of and desire these long-term advantages does not guarantee that the goal will be effectively pursued, any more than the firm desire to live a long life guarantees that a person will give up fatty foods. The hypothesis, then, is that natural selection opted for a special motivational mechanism for this realm: moral conscience. (Joyce, 2006, p. 111)

Note that in the scenario I envision, the soldier has not yet formed strong personal bonds with his fellow soldiers, so that the psychic damage he incurs from a defection strategy can not be attributed to a kin selection mechanism that

has misfired. Also note that the psychic damage is long term, serious, and occurs after the fact to a physically viable organism in its reproductive prime.

Humans and baboons have evolved by natural selection. If you look at the way natural selection works, it seems to follow that anything that has evolved by natural selection should be selfish. Therefore we must expect that when we go and look at the behaviour of baboons, humans, and all other living creatures, we shall find it to be selfish. If we find that our expectation is wrong, if we observe that human behavior is truly altruistic, then we shall be faced with something puzzling, something that needs explaining (Richard Dawkins).

16. I have expressed this hypothesis more colorfully elsewhere:

The reason we turned out like Captain Kirk instead of Mr. Spock (endowed with ego related emotional needs and disorders) or more like Mother Teresa than Joseph Stalin (a evaluative profile that has become “red-shifted” toward evaluative objectivity/ impartiality relative to the “ruthless selfishness” predicted by inclusive fitness theory), has been more a matter of psychodynamic necessity than of adaptive utility (Roberts, 2001).

References

1. Black, Max (1964), ‘The Gap Between “Is” and “Should”’, *The Philosophical Review*, 73/2: 165-91.
2. Cohen, L. Jonathan (1981), ‘Can Human Irrationality Be Experimentally Demonstrated?’, *Behavioral and Brain Sciences*, 4: 317-370; 6: 487-533; 10: 311-13.
3. Dawkins, Richard (1976), *The Selfish Gene*.
4. Drier, James (2001), ‘Humean Doubts about Categorical Imperatives’ in Elio (2001).
5. Elio, Renee (2001), *Varieties of Practical Reasoning*.
6. Godel, Kurt (1962), *On Formally Undecidable Propositions* (translation of 1931 paper).
7. Hamilton, W. D. (1964), ‘The Genetical Evolution of Social Behavior’, *Journal of Theoretical Biology*, 7: 1-16; 17-52.
8. Hume, David (1739), *A Treatise of Human Nature*.
9. Hofstadter, Douglas (1979), *Godel, Escher, Bach: An Eternal Golden Braid*.
10. Joyce, Richard (2006), *The Evolution of Morality*.
11. Lucas, J. R. (1961), ‘Minds, Machines and Godel’, *Philosophy*, 36: 112-27.
12. Penrose, Roger (1994), *Shadows of the Mind*.
13. Penrose, Roger (1989), *The Emperor’s New Mind*.
14. Roberts, Jr., Phil (2005), ‘Feelings of Worthlessness’, first presented before the *Association for Politics and the Life Sciences*, Washington, D. C.
www.rationology.net/abstracts.htm#worthlessness
15. Roberts, Jr., Phil (2006), ‘Rationology 101’, first presented before the *Association for Politics and the Life Sciences*, Bloomington, Ind.
www.rationology.net/abstracts.htm#rationology
16. Roberts, Jr., Phil (2010), ‘Why Human Irrationality Can Not Be Experimentally Demonstrated’, first presented before the *Association for Politics and the Life Sciences*, Bloomington, Ind.
<http://www.rationology.net/abstracts.htm#irrationality>
12. Roberts, Jr., Phil (2001), ‘Why We Turned Out Like Captain Kirk Instead of Mr. Spock’, first presented before the *International Society for the History, Philosophy, and Social Studies of Biology*, Hamden, Conn.
www.rationology.net/abstracts.htm#Kirk
13. Stein, Edward (1996), *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*.
14. Tversky, Amos and Daniel Kahneman (1974), ‘Judgment Under Uncertainty: Heuristics and Biases’, *Science* 185: 1124-31.