

"Heads I win, tails you lose": A Foray Into the Psychology of Philosophy

Tim van Gelder
Department of Philosophy
University of Melbourne
Parkville VIC 3052 Australia
tgelder@ariel.unimelb.edu.au
<http://www.arts.unimelb.edu.au/~tgelder>

One of the classic papers of Australian feminist philosophy is G. Lloyd's "The Man of Reason" (Lloyd, 1979). The main concern of this paper is the alleged *maleness* of the Man of Reason, i.e., the thesis that our philosophical tradition in some deep way associates the concepts *rational* and *male*. Lloyd claims that her main goal is to bring this "undoubted" thesis "into clearer focus" (p.18), and indeed she makes no strenuous effort to demonstrate that the to-be-clared thesis is actually true. There are however a few places where she advances material she seems to be taking as some kind of *evidence* that the Man of Reason is male. One is on the second page, where she quotes from Augustine:

And finally we see man, made in your image and likeness, ruling over all the irrational animals for the very reason that he was made in your image and resembles you, that is because he has the power of reason and understanding. And just as in man's soul there are two forces; one which is dominant because it deliberates and one which obeys because it is subject to such guidance, in the same way in the physical sense, woman has been made for man. In her mind and her rational intelligence she has a nature the equal of man's, but in sex she is physically subject to him in the same way as our natural impulses need to be subjected to the reasoning power of the mind, in order that the actions to which they lead may be inspired by the principles of good conduct.

Now an interesting feature of this passage is that it appears to directly contradict the thesis of the maleness of the Man of Reason. Far from saying that rationality is a male prerogative, Augustine claims that "in her mind and her rational intelligence she has a nature *the equal of man's*" (my emphasis). Certainly Augustine claims that *in sex* woman is subject to man, and he also claims that there is an *analogy* or *parallel* between the dominance of man over woman in sex and the dominance of the rational part of the mind over the natural impulses. However these claims do not, individually or even collectively, imply that there is something distinctively male about rationality or distinctively rational about males, and Augustine goes out of his way to *deny* any difference between men and women in this regard.

Since noticing this anomaly I have conducted a number of informal experiments where the passage is presented to some group, such as a class of undergraduates, who are asked whether Augustine is making an association between maleness and rationality. Overwhelmingly, the answer is *no*. Moreover the participants in these experiments don't seem to have much trouble making up their minds. There is almost no dissent, or perceived need for discussion as if the question or the passage were somehow subtle, difficult or equivocal. That is, to almost all people who come at this

issue "out of the blue," it is *obvious* that Augustine's position contradicts the focal thesis of Lloyd's paper.

I would not go so far as to claim that it is *impossible* to construe Augustine's comments so as to assist Lloyd's case. (Indeed, a few cantankerous colleagues have attempted to do exactly this, though in my opinion their hermeneutical heroics came to nought.) My point, rather, is this. Lloyd is claiming to be exploring an "undoubted" thesis concerning the grand sweep of Western philosophy. If she is right, it ought to be straightforward enough to find relatively clear and convincing examples of philosophers linking rationality and maleness. However, to illustrate her thesis, she selects a passage which most "naive" readers interpret as contradicting her case. So there interesting is an interesting psychological issues here. Why did Lloyd rely on a passage that offers *at best* equivocal support, when less controversial evidence should have been available? Perhaps she genuinely interpreted the passage as providing good support for her main thesis. Does this mean, then, that she somehow simply failed to see what was so obvious to classrooms full of undergraduates? If so, why? And did Lloyd's referees and other pre-publication readers also fail to notice that the Augustine passage, in its most obvious reading, ran counter to the point Lloyd was trying to make? Why have so many of her post-publication readers missed the same point?

I don't know what was actually going through Lloyd's, and for the purposes of this essay it doesn't matter much. Nevertheless, here is a conjecture. The following claims have been maintained by at least some feminist philosophers:

1. Rationality is obviously NOT intrinsically associated with maleness
2. Male philosophers—particularly of the dead white variety—have often taken rationality to be associated with maleness.
3. Since these otherwise intelligent people were denying the obvious, they must have been in the grip of an ideology.

Now, whether or not these claims are true, they can function ideologically, i.e., they can be lenses through which a philosopher perceives and understands the world of philosophy. So the conjecture is that, ironically, Lloyd's prior commitment to the thesis of the maleness of Man of Reason caused her to misinterpret the Augustine passage, just as—supposedly—Augustine's commitment to the maleness of rationality caused him to misinterpret rationality and maleness. Moreover, the passage and its surrounding discussion stayed in the manuscript all the way through to publication because Lloyd's various pre-publication readers were in the grip of the same ideology. (It would, I think, be needlessly cynical to suggest that everyone involved was being just plain negligent.)

Lloyd's apparent blunder is a dramatic example of a pervasive and familiar human failing: roughly, we would rather maul the evidence than change our beliefs. This disturbing phenomenon was never more acutely described than by Francis Bacon, back in the heyday of the Man of Reason:

The human understanding when it has once adopted an opinion draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects or despises, or else by some distinction sets aside and rejects, in

order that by this great and pernicious predetermination the authority of its former conclusion may remain inviolate. (Bacon, 1905)

One might have hoped that philosophers would be exceptions to the general rule. With their years of training in the art of good thinking, philosophers at least should be able to recognize counterevidence and give it proper weight in their deliberations. However Bacon, himself a philosopher, observed that philosophers are, if anything, *especially* liable to distort evidence, falling for a distinctive "Idol of the Mind" (roughly, barrier to good thinking):

Lastly, there are Idols which have immigrated into men's minds from the various dogmas of philosophies, and also from wrong laws of demonstration. These I call Idols of the Theater, because in my judgment all the received systems are but so many stage plays, representing worlds of their own creation after an unreal and scenic fashion.

Anyone who reads the Augustine passage as associating maleness and rationality has fallen under the spell of an Idol of the Theater, viz., the thesis of the Maleness of the Man of Reason.

Irrational Persistence of Belief

Over the years many authors have claimed that humans tend to be excessively protective of their own beliefs, and there are any number of anecdotal illustrations. Unfortunately, these scenarios are often tragic; to take an extreme case, historian Hugh Trevor-Roper has claimed that "there is no record of Hitler's ever having changed his mind, once it was made up," a rather extreme case of cognitive intransigence. More mundane examples are easily found in our own experience. Indeed, every time we fall into a heated dispute, our opponent is obviously failing to accord due weight to overwhelming counterevidence!

Naturally such a salient and pervasive phenomenon has attracted considerable attention from professional psychologists. In the past few decades they have conducted dozens of careful studies of what they sometimes call "belief preservation" and sometimes call "irrational persistence of belief." They have teased out a number of subtly different aspects of the general phenomenon, and conducted a number of ingenious experiments generating both quantitative and qualitative insight.

Sometimes we believe a claim, whether we want it to be true or not; and sometimes we want a claim to be true, whether we believe it or not. Often we do both, and often it is hard to distinguish the two—as when, for example, somebody claims to sincerely believe that Jesus is their friend. For most of what follows it doesn't matter much whether we talk of believing or wanting-to-be-true; both attitudes give rise to belief preservation phenomena. For this reason I will use the term "pro-attitude" to cover both kinds of attitude, either separately or in combination.

In an important textbook, Jonathon Baron defines irrational persistence of belief as

the tendency to search for evidence, and to use evidence, in a way that supports beliefs that are already strong for us (prior beliefs) or beliefs that we want to be true (Baron, 1994 p. 280)

and claims that it is "one of the major sources of human folly" (p.281). He immediately points out, however, that in any given case demonstrating that someone is *irrationally* maintaining belief can be rather difficult, for we have no simple algorithms for determining when belief is properly proportioned to evidence. Whose golden epistemic standard fixes the bounds of the rational? As Baron remarks, the great scientist Michael Faraday persisted in believing that electrical currents could be induced with magnets, despite counterevidence in the form a number of failed attempts to induce such currents experimentally. It would be difficult to maintain that he was just being irrational.

To get around this problem, psychologists avoid taking sides on the issues themselves, and focus instead on investigating whether and to what extent people violate certain very general, *topic-neutral* norms or principles of good reasoning. For example, I cannot simply rule that belief in a Christian god is irrational due to overwhelming counterevidence, for attempting to justify that claim leads directly into the tangled thickets of theological debate. However I can confidently say that anyone whose belief in a Christian God is *strengthened when presented with counterevidence* is being irrational.

Following this approach, psychologists have identified at least five major respects in which people tend to exhibit irrational belief persistence:

1. A pro-attitude makes us seek supporting evidence and ignore conflicting evidence.

One of the most obvious manifestations of belief preservation arises when we consider whether a claim merits our acceptance. When we have a pro-attitude to the claim we tend to *actively seek* evidence confirming or supporting the claim, and *fail to seek* evidence going against it. That is, in our search for evidence we try to bolster our beliefs rather than challenge them. For an obvious example, people inclined to believe that Jewish people are especially avaricious note any apparent instance of this alleged trait with glee, and show no interest in seeking out counterexamples. As J.S. Mill noted, this tendency rules searches in memory as well as searches for new information:

It is evident that when the instances on one side of a question are more likely to be remembered and recorded than those on the other, especially if there be any strong motive to preserve the memory of the first, but not of the latter, these last are likely to be overlooked, and escape the observations of the great mass of mankind.

One study designed illustrating this phenomenon was conducted during the 1964 U.S. presidential campaign (Lowin, 1967). Subjects were provided with samples of brochures supporting various candidates, and were then able to request more brochures. When the samples contained strong and persuasive arguments, subjects ordered a greater number of brochures supporting their own candidate, thereby bolstering their existing attitude. Interestingly, when the arguments were weak and easy to refute, subjects ordered more brochures supporting the opponent. It seems they were strengthening their belief by refuting the arguments on the other side, something they avoided if those arguments were not easily knocked down.

2. A pro-attitude to a claim makes us more likely to judge an argument for that claim to be valid.

We are all familiar with the situation where our opponent in some dispute appears to accept uncritically any argument which appears to support their position, no matter how dismal, while at the same time rejecting compelling arguments against it. In this vein, psychologists have shown that people are more likely to accept an inference as logically valid if they already believe the conclusion, and more likely to reject an inference as invalid if they already believe the conclusion to be false.

For example, Evans, Barston & Pollard (1983) presented subjects with syllogisms of the form

No cigarettes are inexpensive.
Some addictive things are inexpensive.
Therefore, some addictive things are not cigarettes.

and

No addictive things are inexpensive.
Some cigarettes are inexpensive.
Therefore, some cigarettes are not addictive.

These are both valid syllogisms of the same form; the key difference is the plausibility or otherwise of the conclusion. Subjects nearly always judged the syllogisms with a plausible conclusion as valid, but a majority of the time judged the syllogism invalid if it had a false conclusion.

3. A pro-attitude makes us overweigh supporting evidence and underweigh conflicting evidence.

This effect is even more pronounced in situations involving informal argumentation, in which a conclusion ought to be accepted or rejected on the balance of considerations, and no calculus determines the appropriate response. In such situations we are forced to intuitively assess the overall weight and direction of the evidence, and intuitive assessments are especially prone to contamination by pro-attitudes.

One way psychologists study this topic without taking any sides on the issues involved is to measure the extent to which people adhere to, or violate, the so-called "neutral evidence" principle. According to this principle, your confidence in a claim should not be increased by evidence that is neutral in the sense that it offers equal support to your position and its opposite. In fact, people tend to treat neutral evidence as, on balance, increased the support for their favoured position.

This tendency was dramatically illustrated in a study in which subjects who either favoured or opposed capital punishment were given sets of concocted journal reports describing studies of whether capital punishment deters murder (Lord, Ross, & Lepper, 1979). Some reports indicated that capital punishment does deter murder; other reports indicated that it does not. Some reports compared murder rates in states with or without capital punishment; others considered data from states before and after the introduction of capital punishment. Each subject was given two reports, and careful "mixing and matching" ensured that each subject received an appropriate selection from each category.

When evidence is genuinely *mixed*—some supporting our position and an equal amount going against—it should *reduce* our confidence. As Baron points out, "If people have stronger (or more) evidence for the side they favor, then mixed evidence, which is equally strong on both sides, would add proportionately more strength to the other side" (p.288). If people treated evidence in a self-serving biased way, we would expect them to fail to reduce their confidence sufficiently. In fact, people can be much more irrational than that. In the Lord *et al* study subjects *polarised*: their initial belief became stronger, regardless of whether it was for or against capital punishment. Further, subjects rated the report which agreed with their opinion as "more convincing" and found flaws more easily in the report that went against their opinion.

Baron dryly notes: "This study is disturbing, because it suggests that evidence is useless in settling controversial social questions" (p.288)—a conclusion one might well have reached independently, observing current debates such as whether certain drugs should be legalised.

4. A pro-attitude can result in negative evidence having a reinforcing effect.

It is, perhaps, not all that surprising that neutral evidence tends to be interpreted as weighing in behind one's pro-attitudes. Rather more astonishing is the fact that negative evidence *alone* can result in strengthened belief. As David Stove put it, we often treat evidence on a "heads I win, tails you lose" basis. This phenomenon was illustrated in a classic study by Batson (1975), who presented a group of believers in Christ's divinity with authoritative-seeming evidence purporting to demonstrate, on the basis of newly discovered scrolls, that the New Testament was a fraud. Those who accepted the veracity of the evidence ended up *even more* convinced of Christ's divinity than before. As Baron put it, "The believers who accepted the article had the greatest need to strengthen their belief in the divinity of Jesus, and they did so despite being given nothing but negative evidence" (p. 289).

This phenomenon was already noted by Mill in *The Subjection of Women*:

So long as an opinion is strongly rooted in the feelings, it gains rather than loses in stability by having a preponderating weight of argument against it. For if it were accepted as a result of argument, the refutation of the argument might shake the solidity of the conviction; but when it rests solely on feeling, the worse it fares in argumentative contest, the more persuaded its adherents are that their feelings must have some deeper ground, which the arguments do not reach; and while the feeling remains, it is always throwing up fresh entrenchments of argument to repair any breach made in the old.

Anecdotal evidence suggests that religious belief is a particularly fertile field for this kind of strengthening in the face of counterevidence. For example in 1997 a Kansas woman claimed that her Virgin Mary plaque weeps blood, which miracle of course brought hordes of visitors. DNA tests commissioned by the local Catholic diocese found that the blood was that of the plaque owner. To most of us, this suggests that God was not actually involved. However the editor of the local newspaper took a different approach, inferring that the event was *even more* miraculous, for obviously God had managed to create blood exactly matching that of the owner.

One is reminded of J.K. Galbraith's incisive observation: "Faced with the choice between changing one's mind and proving that there is no need to do so, almost everyone gets busy on the proof."

5. A pro-attitude can lead to persistence after total discrediting.

A final dimension worth mentioning is what has come to be known as persistence after total discrediting. It occasionally happens that we arrive at a belief on reasonable grounds, but those grounds are later removed. The ideally rational person would then surrender the belief (unless, of course, alternative grounds have since been found). Often, however, people maintain the belief despite the total discrediting of their available evidence. For example, Anderson *et al.* gave one group of subjects evidence that that firefighting and risk-taking are positively associated, and another group evidence that they are negatively associated. Subjects were then told that the evidence was totally made up. When asked afterwards, they showed a strong tendency to believe what they were first told, even after becoming aware that their evidence was worthless (Anderson, Lepper, & Ross, 1980).

In short, beliefs pick up a kind of epistemic momentum: once in motion, they stay that way even when the original evidential forces are removed.

Conclusion

As best we can tell, irrational persistence of belief affects everyone to some greater or lesser degree. As T.C. Chamberlin noted last century, even trained scientists are not immune to its pernicious effects. In his delightfully old-fashioned metaphor, scientists tend to treat their favoured theories with parental partiality:

Love was long since represented as blind, and what is true in the personal realm is measurably true in the intellectual realm....

There is an unconscious selection and magnifying of the phenomena that fall into harmony with the theory and support it, and an unconscious neglect of those that fail of coincidence. The mind lingers with pleasure upon the facts that fall happily into the embrace of the theory, and feels a natural coldness toward those that seem refractory. Instinctively there is a special searching-out of phenomena that support it, for the mind is led by its desires. There springs up, also, an unconscious pressing of the theory to make it fit the facts, and a pressing of the facts to make them fit the theory. When these biasing tendencies set in, the mind rapidly degenerates into the partiality of paternalism.... (Chamberlin, 1965)

Philosophers, of course, are only human, and until somebody provides good evidence otherwise, we should assume that philosophers are innately prone to irrational belief preservation just like everyone else. In professional practice, this would manifest itself as tendencies to devote far more energy to arguing *for* one's position than to *arguing* against it; to rate arguments supporting one's position as more decisive than those opposing it; and to maintain or even increase one's confidence in a position even when the balance of argument goes the other way. In short, we should expect certain kinds of systematic irrationality to pervade philosophical debate as much as any other domain of human thought.

There is however an important difference between philosophers and anyone else, including scientists. After all, epistemology is a branch of philosophy, and philosophers pride themselves on being especially self-conscious about the nature of good thinking. Thus philosophers more than anyone else ought to be aware of the nature and pernicious effects of irrational belief preservation, and to take measures to counteract it. The fact that almost all philosophers are almost entirely *oblivious* to the psychological phenomenon is, I would say, a serious professional delinquency. Francis Bacon understood that

The mind of man is far from the nature of a clear and equal glass, wherein the beams of things should reflect according to their true incidence; nay, it is rather like an enchanted glass, full of superstition and imposture, if it be not delivered and reduced.

Psychologists have studied the flaws in the glass in considerable detail, with obvious implications for how it might be "delivered and reduced." Typically, however, philosophers blithely ignore these insights, preferring to rely on a dubious mix of archaic theories and their own unreliable intuitions.

Bertrand Russell once claimed that "To avoid the various foolish opinions to which mankind are prone, no superhuman genius is required. A few simple rules will keep you, not from *all* error, but from silly error." (Russell, 1950 p.135) In the spirit of Russell's suggestion, here are a few simple methodological principles which, in light of the available empirical evidence on belief preservation, all philosophers would do well to observe in their professional activities:

1. Maintain an active awareness of the psychological phenomenon of irrational persistence of belief.
2. Consciously identify the positions towards which one has a pro-attitude.
3. Be extra diligent in seeking arguments against those positions.
4. Be extra diligent in seeking weaknesses in the arguments for that position.
5. Assign extra credit to arguments for the opposing position.

Following these rules does not mean discarding your beliefs just because they are yours, or surrendering them as soon as the slightest counterevidence arises. It does mean compensating carefully for known biases, much as a skilled archer will adjust his aim to allow for the influence of a breeze.

References

- Anderson, C. A., Lepper, M. R., & Ross, L. (1980) Perverserance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, **39**, 1037-1049.
- Bacon, F. (1905 (1620)) *Novum Organum* (Ellis, R. & Spedding, J., Trans.). London: Routledge.
- Baron, J. (1994) *Thinking and Deciding* (2nd ed.). Cambridge: Cambridge University Press.

- Batson, C. D. (1975) Rational processing or rationalization? The effect of disconfirming evidence on a stated religious belief. *Journal of Personality and Social Psychology*, **32**, 176-184.
- Chamberlin, T. C. (1965) The method of multiple working hypotheses. *Science*, **148**, 754-759; reprinted from *Science*, 1890.
- Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983) On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, **11**, 295-306.
- Lloyd, G. (1979) The Man of Reason. *Metaphilosophy*, **10**, 18-37.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, **6**, 1-9.
- Lowin, A. (1967) Approach and avoidance: Alternative modes of selective exposure to information. *Journal of Personality and Social Psychology*, **6**, 1-9.
- Russell, B. (1950) *Unpopular Essays*. London: George Allen and Unwin Ltd.