

accordingly engage in self-deception when its beneficial interpersonal consequences are most substantial? If so, through what means do individuals “know” that self-deception is likely to be effective? More generally, how has selection shaped the processes that regulate self-deceptive tactics in ways that, ancestrally, enhanced net fitness benefits? Though self-deceptive processes have been studied by psychologists for decades, we will lack a deep understanding of them until researchers take to heart VH&T’s fundamental point.

I take up here one component of self-deception’s functional context that VH&T are surely aware of but do not explicitly discuss in any detail. In the interpersonal context in which self-deception operates, other-deception via self-deception may be in the interest of the actor, but *not* being deceived by others is typically in the interest of all target perceivers. That is, selection should favor perceivers whose inferences are not readily manipulated in the interests of others. Hence, self-deception not only must be understood in an evolutionary framework, but it also must be appreciated in the context of a *co-evolutionary* framework in which the targets of other-deceit must be assumed to have been subject to selection to avoid being deceived (e.g., Rice & Holland 1997).

In the 30-plus years since the concept of self-deception was introduced, biologists have developed sophisticated theories pertaining to communication between organisms, now collectively referred to as signaling theory (e.g., Searcy & Nowicki 2005). One core component of signaling theory is the concept of honest signaling. The idea is that, for any signaling system to evolve, both senders and receivers must benefit, and for receivers to benefit, the signal must contain accurate or “honest” information. If the signal is dishonest (e.g., the size of peacocks’ tails reveal nothing about the quality of their bearers), receivers should evolve to ignore it – that is, the signaling system should collapse (or fail to evolve in the first place).

In light of this notion, how have self-deceptive processes aimed to deceive others been maintained by selection? If perceivers suffer from attending to performances rendered deceptive by self-deception, why has selection not led them to ignore such performances? Perhaps most notably, why should perceivers be fooled by false confidence bolstered by self-deception? Several possibilities come to mind.

First, most performances may well be honest portrayals of earned or honest confidence. A signaling system that is basically honest (honest on average) can tolerate some level of dishonesty (e.g., Searcy & Nowicki 2005).

Second, in many circumstances it may be difficult for individuals to detect the difference between a performance backed by earned confidence and one enabled by self-deceived confidence. An implication of the co-evolutionary nature of signaling systems is that perceivers should be attentive to cues of false confidence – for instance, through utilization of multiple cues and reliance on reputation based on past performance as well as current performance. They should furthermore not tolerate false confidence even when the actor is unaware of its nature. (One reason why narcissistic individuals have unstable interpersonal relations is because their unearned arrogance leads others to reject them.) Yet one may be able to successfully perform with false confidence under conditions in which other information is lacking (such as one-shot interactions). This possibility once again underscores the need for researchers to examine the contexts in which self-deception affects performance, including the audiences of those performances in light of their desire not to be fooled.

Third, confidence bolstered by self-deception may not, ultimately, be other-deceptive. Individuals who have earned confidence may nonetheless benefit by carrying off that confidence by self-deceiving (e.g., not attending to their own shortcomings). In this view, individuals who can best afford to self-enhance through self-deception are those who are viewed positively by others in any case. Ironically, in this view, self-deception facilitates honest, not deceptive, social performance.

These possibilities are not mutually exclusive. And there may be others.

More generally, the fact that social performances enhanced by self-deception must be understood in the context of co-evolved audience resistance to falsely enhanced performance has implications for how self-deception should be studied. And beyond that, at a basic theoretical level, it suggests ways in which an appreciation for the evolutionary processes that have shaped self-deception should be deepened.

Culture of deception

doi:10.1017/S0140525X10003122

Gregory Gorelik^a and Todd K. Shackelford^b

^aDepartment of Psychology, Florida Atlantic University, Boca Raton, FL 33431;

^bDepartment of Psychology, Oakland University, Rochester, MI 48309.

ggorelik@fau.edu

shackelf@oakland.edu

<http://www.toddkshackelford.com/>

Abstract: We examine the self-deceptive aspects of religion and nationalism. By embracing various religious or political ideals, regardless of their truth, our ancestors could have enhanced their confidence, solidified their social ties, and manipulated their reproductive rivals. This use of culture as one’s extended phenotype may increase the spread of misinformation and create global webs of deception and self-deception.

If humans have evolved a capacity to deceive themselves so as to better deceive others, then human technologies, languages, ideas, and traditions might display cultural manifestations of deceptive and self-deceptive adaptations. Deceiving oneself may be easier if others are complicit in the deception. Collective self-deception is manifested as groupthink and deindividuation, and it is likely mediated and enabled by various cultural elements. Von Hippel & Trivers (VH&T) briefly discuss the social reinforcement of individual-level self-deception, but they do not elaborate upon the full implications of the cultural aspects of self-deception. We discuss the ways in which self-deception may be expressed collectively in religious and political contexts, and we present several possibilities for how gene-culture co-evolution has affected human deception and self-deception.

According to Dawkins’s (1982) concept of the extended phenotype, genes are selected for how well they code for an organism’s ability to manipulate its environment. An organism’s environment includes other organisms, of both the same and different species. Therefore, organisms may be selected for how well they can manipulate other organisms, effectively using them as extended phenotypes of their own selfish genes. If humans have competed with one another over reproductively relevant resources throughout their evolutionary history, then selection pressures may have sculpted adaptations by which humans manipulate and deceive their reproductive rivals. In addition, given the human capacity for non-genetic transfer of information (i.e., culture), many cultural phenomena may display design features indicative of their use in deceiving oneself and others. Therefore, human genes may be selected for how well they code for psychological programs that use cultural information to deceive other humans. In effect, culture is part of our extended phenotype and is an integral part of the environment to which our genes have evolved.

Following this line of thought, we can investigate human culture for features that enable its use during deception of oneself and others. Organized religion and nationalism display several exemplar features. In most ancestral contexts, religious or political self-deception may have benefited individual

members, but there was a risk of exploitation if some individuals accepted the benefits of membership without paying the costs of helping other members. In such instances, the institution in question could have been used as a tool by which some individuals manipulated others. If manipulators benefited by their manipulation, then manipulative traits may have proliferated throughout human populations (until the costs of manipulation outweighed the benefits). At the same time, the cultural tools that manipulators used to express their manipulative traits might have been refined and passed down the generations alongside the genetically coded, manipulative psychological programs. In this way, genes and culture depend on each other for the evolution and expression of deceptive and self-deceptive adaptations.

Various design features of religious and political institutions may be indicative of their role in deception and self-deception. As described by VH&T (sect. 5.5.2, para. 1), insecure societies display higher rates of religious belief, because belief in God may provide individuals with a sense of control over their lives. Assuming that this sense of control was advantageous for our ancestors because it enabled the manipulation of reproductive rivals, it should then be no surprise that humans are willing and able to accept as true certain fantastic doctrines and dogmas. Likewise, religion and nationalism exhibit a strength-in-numbers effect that facilitates collective self-deception. The costs of religious or political misinformation may not offset the benefits of joining and supporting such institutions. Therefore, the deception of individual members is made easier by the pervasiveness of self-deception within these institutions.

There are other features of organized religion and nationalism that portray self-deceptive qualities. The avoidance of information that threatens or could weaken a religious or political institution is ubiquitous. This is seen when totalitarian regimes limit the types of media that are available to the public, or when religious followers avoid being exposed to competing doctrines or scientific facts (i.e., evolution by natural selection). If exposed to threatening information, followers may attempt to rationalize away whatever threat they were exposed to or be skeptical of this information. In this way, patriots from one nation may doubt the veracity of a rival nation's messages and ideas by calling them propaganda. Likewise, creationists sometimes tie themselves into psychological knots in attempting to explain away the evidence for evolution (when they do not deny or ignore this evidence altogether).

Derogation of others and enhancement of oneself are also common features of nationalism and religion. Some examples of this include the American motto "one nation, under God," or the belief that one is a member of the "chosen people" or of the "master race," while dehumanizing members of other nations or religions. Furthermore, optimism about the future is pervasive within religious and political circles. This optimism can lead to a self-fulfilling prophecy if one is motivated to action by the promise of a political utopia or a heavenly paradise, but it also can be used to manipulate members into acting against their own interests. Likewise, such cultural modes of self-enhancement may increase one's confidence and lead to social solidarity with one's community, but they also may bring about social conflict and war.

According to VH&T, convincing oneself that a lie is true while knowing that it is false at some psychological level is the most extreme form of self-deception. Religion, in particular, may use the consequent cognitive dissonance to its advantage by pointing to this internal conflict as evidence of its veracity. The constant struggles to retain one's faith or to remain spiritual amid the onslaught of secularism seem to be essential features of modern Judaeo-Christian practices. In this way, religion may be an especially useful cultural tool by which individuals manipulate their rivals by imposing self-deception upon them.

Deceiving ourselves about self-deception

doi:10.1017/S0140525X1000227X

Stevan Harnad

Institut des sciences cognitives, Université du Québec à Montréal, Montreal, QC H3C 3P8, Canada; School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, United Kingdom.

harnad@ecs.soton.ac.uk

<http://www.ecs.soton.ac.uk/people/harnad>

Abstract: Were we just the Darwinian adaptive survival/reproduction machines von Hippel & Trivers invoke to explain us, the self-deception problem would not only be simpler, but also nonexistent. Why would *unconscious* robots bother to misinform *themselves* so as to misinform others more effectively? But as we are indeed conscious rather than unconscious robots, the problem is explaining the causal role of consciousness itself, not just its supererogatory tendency to misinform itself so as to misinform (or perform) better.

Von Hippel & Trivers (VH&T) are deceiving themselves – with the help of adaptivist psychodynamics and a Darwinian Unconscious. They have not proposed an adaptive function for self-deception; they have merely clad adaptive interpersonal behaviour in a non-explanatory mentalistic interpretation: *I can persuade you more convincingly that I am unafraid of you (or better fool you into thinking that the treasure is on the right rather than the left, where it really is) if I am unaware of – or “forget” – my own fear (or the fact that the treasure is really on the left rather than the right).*

Sure. But then in what sense am I afraid at all (or aware where the treasure really is)? If I *feel* (hence act) afraid, then you detect it. If I don't feel the fear (or the sinistroversive urge), then I don't act afraid, and you don't detect any fear (because there is nothing there to detect).

So in what sense am I “self-deceived”? (Ditto for left/right.) Is it always self-deception not to feel afraid (or not to remember that the treasure's on the right), when I “ought to” (or used to)?

The same is true of “self-enhancement”: Yes, I am more convincing to others, hence more influential on their behaviour, if I behave as if I expect to succeed (even when I have no objective grounds for the expectation). But in what sense am I self-deceived? In feeling brave and confident, when I “ought to” be feeling fearful and pessimistic? Shouldn't organisms all simply be behaving in such a way as to maximize their adaptive chances?

In fact, what does what organisms *feel* have to do with any of this at all (apart from the entirely unexplained fact that they do indeed feel, that their feelings are indeed correlated with their adaptive behaviour, and that their feelings do indeed feel causal to them)? The feelings themselves (i.e., consciousness) are much harder to situate in the adaptive causal explanation – unless you believe in telekinesis (Harnad 2000)! (Hence, I feel that VH&T have bitten off a lot more here, phenomenally, than they can ever hope to chew, functionally.)

The treasure is the best example of all, because that is about *facts* (data) rather than just feelings: Suppose I did indeed “know” at some point that the treasure was on the left – in the sense that if at that point I could have reached for it without risk of being attacked by you, I would have reached for it on the left. But, according to VH&T, it was adaptive for me to “forget” where the treasure really was, and to believe (and behave as if) it was on the right rather than the left, so as to deceive you into heading off to the right so I could eventually grab the treasure on the left and dart off with it.

But isn't the true adaptive design problem for the Blind Watchmaker – apart from the untouched problem of how and why we feel at all (Harnad 1995) – a lot simpler here than we are making it out to be (Harnad 2002)? And are we not deceiving ourselves when we “adapt” the adaptive explanation so as to square with our subjective experience?