Chapter 18
Deceit and Self-Deception

Robert Trivers

Abstract  Deception is a universal feature of life, at all levels and in all relationships both within species and between species, inside individuals and outside, with strong effects on both deceiver and deceived. Being detected often results in a sharp reversal of fortune for the deceiver thereby intensifying selection to deceive successfully. In encounters between human strangers, nervousness, signs of control and of cognitive load can all serve as cues of deception but cognitive load appears to be the most important. Self-deception is defined as hiding true information from the conscious mind in the unconscious, and is illustrated by classical experimental work. Selection to deceive can favor self-deception, the better to hide the deception and separately to reduce its cognitive costs. Four examples are described. There is a general tendency toward self-inflation in humans, the better to give off a positive image. Conscious thought suppression, studied via fMRI, shows that one area of the brain has been coopted to suppress memory formation elsewhere in the brain. When people reach age 60, they fail to attend to negative social reality, and this old-age positivity may give immune benefits. Across primates there is a strong positive association between relative size of the neocortex and frequency of deceptive acts in nature. If the relationship holds within species, we may expect relatively intelligent humans to be prone to self-deception. There is such a thing as imposed self-deception, in which we act out the system of self-deception of another. Likewise, there is parasitized self-deception in which our system of self-deception makes us more vulnerable to deception by others. Con artists are given as an example. One could model the evolution of deceit and self-deception as a multiplayer game, which can then be analyzed mathematically, modeled via computer simulations or tested experimentally. One promising possibility is a variant of the Ultimatum Game, in which deception and detection of deception are permitted and given quantitative values.

R. Trivers
Department of Anthropology, Rutgers University, Rutgers, 131 George St, NJ, USA
e-mail: trivers@rci.rutgers.edu

P.M. Kappeler and J.B. Silk (eds.), Mind the Gap,
DOI 10.1007/978-3-642-02725-3_18, © Springer-Verlag Berlin Heidelberg 2010
18.1 Introduction

In our own species, deceit and self-deception are often two sides of the same coin - best seen together. If by deception, we limit ourselves to consciously propagated deception - lies - we will miss the larger category of unconscious deception, including active self-deception. On the other hand, if we look at self-deception and fail to see its connection to deception, we will miss its major function. The central claim is that self-deception evolves in the service of deception (1) to avoid its detection and (2) to reduce the immediate cognitive costs. In the first case, the self-deceived person fails to give off the cues that go with consciously mediated deception, thus escaping detection. In the second, the actual process of deception is rendered less expensive by keeping part of the truth in the unconscious mind. But this turns out to be a very complicated matter. Suppressing the truth may have short-term cognitive benefits as well as almost immediate immunological costs, or for that matter, longer-term cognitive ones. In addition, deception and self-deception are also being directed at you by others, so that there is such a thing as imposed self-deception, in which you act out the system of self-deception of another person, presumably often to your own detriment. In this account, we will try to fuse the two topics of deceit and self-deception into a single coherent account, with special attention to the associated costs and benefits, that is, with their selective effects.

18.2 Deception is Everywhere, at All Levels of Life

Deception (misleading others) is a very deep feature of life. It occurs at all levels and, it would seem, by any means possible. It tends to hide from view and is, by its nature, difficult to see and to study. Self-deception is even worse, hiding itself more deeply in our own unconscious minds. It is, thus, important to build up our logic carefully, with a full view of the enormous variation that has evolved for various reasons (Trivers 2000). Unfortunately, there has been very little formal theory on the subject, so I will limit myself to making a series of general comments, with special focus on cognitive load and self-deception.

When I say that deception occurs at all levels of life, I mean that viruses practice it, bacteria do, so do plants, and so do the insects preying on plants, and a wide range of other animals. It is everywhere. Even within our genomes deception may flourish, as selfish genetic elements use deceptive molecular techniques to over-reproduce at the expense of the larger genome (Burt and Trivers 2006). Or, when a selfish paternal orientation collides with an oppositely oriented maternal one (Haig 2002). Deception infects all the fundamental relationships in life, parasite and host, predator and prey, plant and animal, male and female, neighbor and neighbor, parent and offspring (including mother and fetus), and even the relationship of an organism to itself.

Viruses and bacteria actively deceive to gain entry in to their hosts, by mimicking body parts for example. Or, as in HIV, the virus deceives by changing coat proteins so often as to make mounting an enduring defense almost impossible. Predators gain from being invisible to their prey or resembling items attractive to them while prey gain by being invisible to their predators or mimicking items noxious to them, e.g., poisonous species or their predator’s predator.

Deception within species requires only imperfect (<1) degrees of relatedness between individuals (so that their self-interests are not identical) and imperfect information (so that the other party can be fooled). Both are easily satisfied. Clonal species are rare in nature and imperfect degrees of relatedness are the rule (under out-breeding, 1/2, 1/4 down to near-zero). Perfect information is impossible. Deception can allow you to steal or induce the transfer of food and other resources, engage in extra-pair copulations undetected, manipulate your parents, your mate, your offspring, your neighbors - even the maternal (or paternal) half of yourself.

And deception always takes the lead, while detection of deception plays catch-up. As has been said regarding rumor, the lie is halfway around the world, before the truth goes to work. When a new deception shows up, it starts rare in a world that lacks a proper defense. As it increases in frequency, it selects for such defenses in the victim, so that eventually its spread will be halted by the appearance and spread of countermeasures, but new defenses can always be bypassed and new tricks invented.

The adaptive potential of deception is chronically overlooked by those with an attachment to the truth. Bill Gates told the world confidently, in 2004, that the problem of spam “would be solved by 2006” (N.Y. Times 12/06/06). He saw that defenses could easily be erected against the set of spamming devices then in use but he could not imagine that these defenses could easily be bypassed while yet newer forms of spamming were continually being invented. Spam is now at least ten-fold as frequent as it was in 2004, and correspondingly costly. One inevitable cost is the destruction of true information by spam detectors too stringently set, including by oneself. This is a universal problem in animal discrimination. Greater powers of discrimination will inevitably increase the so-called false negatives, rejecting something as false which is, in fact, true.

It always amuses me to hear economists saying that the costs of deceptive excesses in our economy (so-called “white” crime robbery) will be naturally checked by “market forces.” Are these the same forces that force us to add one unit to every price we see in order to know the true price? Why should the human species be immune to the general rule that where selection is strong, deception can be generated that extracts a substantial net cost every generation. Consider the following.

Deception is such an important feature of life that it can entrain the evolution of entire groups of organisms, as well as the evolution of specialized deceitful morphs within species. For example, the Phasmatodea, or stick insects, is a group that has given itself over to the imitation of either sticks (~3,000 species) or leaves (~30 species) (Markle 2007). In the case of sticks, there is apparently a tremendous evolutionary pressure to produce a long, thin (stick-like) body, even if this forces...
the individual to forego the benefits of bilateral symmetry. Thus, to fit the internal organs into a diminishing space, one of two organs have often been sacrificed, only one kidney, one ovary, one testis, etc. This shows that selection for successful deception has been powerful enough not only to remodel the external shape of the creature but also to remodel its internal organs as well — even when this is otherwise disadvantageous to the larger organism, as loss of symmetry, in principle, must usually be.

Likewise, selection for deception has been strong enough to mold morphs that are obligately committed to deception; that is, morphological forms whose strategy depends entirely on deception of others. A classic example occurs in the blue-gill sunfish, where a specialized male form has evolved that mimics a female in appearance and behavior, being 1/6th the size of a territorial male and roughly the size of an actual female (for a recent reference see Stoltz and Neff 2006). This female-mimic seeks out a territorial male, permits himself to be courted, and responds enough to keep the other male interested, so that when a true female spawns the pseudo-female is ready nearby, along with the territorial male, to fertilize the eggs. It is as if the territorial male imagines he is in bed with two females when, in fact, he is in bed with one female and one male. The two kinds of males appear to be distinct forms that never turn into each other. To have persisted for so long their long-term reproductive success must be identical; that is, the deceiver is doing exactly as good as the deceived — and this equality must, in turn, have been produced by frequency-dependent selection. That is, when the female-mimic is relatively rare, he will do relatively well; when common, less so.

18.3 Detection of Deception Often Leads to Negative Consequences, Including Punishment

In our own species, we hardly need convincing that when our deception is detected, we may receive some harsh feedback, a beating, a public humiliation, a lost love, or at the very least, a withdrawal of some trust and affection. This appears to be true of many other animals. That is, in insects and birds, deceivers are often punished aggressively, and most often by those whose status is being threatened by the deceiver (for some examples see Trivers 1985; Möller 1987; Hauser 1992; Tibbetts and Dale 2004). If we take it as generally true that detection of deception often leads to an unfortunate reversal in fortune, then a deceiver will be under pressure when deceiving and these signs of pressure are available for detection by others. Indeed, the lie detector test is based on this fact.

But here a caution is worth mentioning. The lie detector test continuously measures a series of physiological parameters in response to a series of posed questions designed to detect unusual signs of stress in particular places. For example, the most useful device is the “guilty knowledge” test, which suddenly introduces some pertinent fact known only to the guilty (was the gun red?). Here, any deviation from the norm — high reactivity or suppressed reactivity — shows guilty knowledge.

Daily life is very different from this. At one extreme, such as intimate relations between husband and wife, or partner and partner, each may indeed have a detailed behavioral template of the other, honest and dishonest, against which to evaluate the ongoing behavior, while at the other extreme, perfect strangers are interacting with no prior knowledge of each other.

18.4 Cognitive Load is a Key Factor in the Detection of Deception in Humans

In anonymous or infrequent interactions, behavioral cues can not be read against a background of known behavior, so more general attributes of lying must be used. Three such attributes have been emphasized (Vrij 2004):

1. **Nervousness**: because of the negative consequences of being detected, including being aggressed against, nervousness can reveal deception (DePaulo et al. 2003)
2. **Control**: in response to concern over appearing nervous (or concentrating too hard), people exert control, trying to suppress both, but this requires additional effort, with possible side effects, and there is the danger of over-acting, over-control, a planned or rehearsed impression
3. **Cognitive load**: lying is cognitively demanding; you must suppress the truth and construct a falsehood, one that will not contradict anything known by the listener, or likely to be known, you must tell it in a convincing manner and in such a way that you can remember the story. This takes time and concentration, both of which may give off secondary cues.

The most recent work suggests that cognitive load is the critical variable among the three, with a minor role for control and very little for nervousness. At least this seems to be true in real criminal investigations as well as experimental situations designed to mimic them (Mann and Vrij 2006). People who are lying have to think too hard and this causes several effects, some of which are opposite to those of nervousness.

Consider, for example, blinking. We blink our eyes more often when nervous, but we blink them less under increasing cognitive load (e.g., while solving arithmetic problems). Recent studies of deception suggest that we blink less when deceiving, that is, cognitive load rules (Vrij 2004; Vrij et al. 2006). Nervousness makes us fidget more but cognitive load has the opposite effect. Again, contra usual expectation, people fidget less in certain deceptive situations. Again consistent with cognitive load effects, men use less hand gestures while deceiving and both sexes show longer pauses while they speak.

The effects of control are illustrated by pitch of voice and (separately) displacement activities. Deceivers tend to have higher-pitched voices (DePaulo et al. 2003).
This is a natural consequence of any effort to suppress behavior by becoming more rigid. Tensing up the body inevitably raises the pitch of voice.

Another effect of suppression is the production of displacement activities. As classically described in other animals, these are irrelevant activities often seen when two opposing motivations are simultaneously aroused. Since neither impulse can express itself, the blocked energy easily activates irrelevant behavior, such as a twitch. For this reason, displacement activities in primates are a reliable indicator of stress (Troisi 2002).

Nervousness is almost universally cited as a factor associated with deception both by those trying to detect deception as well as by those trying to avoid detection, yet surprisingly enough, it is one of the weaker factors predicting deception. This is, perhaps, precisely because we are conscious of our nervousness so that mechanisms of suppression may be almost as well developed as the nervousness itself. The point about cognitive load is that there is no escape. If it is cognitively expensive to lie, there is no obvious way to reduce the expense. Mechanisms of denial and repression may serve to reduce immediate cost, but alas, as we shall see, with ramifying costs later on.

18.5 What is Self-Deception?

Before going any further it would be useful to define self-deception, and give an example. Some people, especially philosophers, imagine that self-deception is a contradiction in terms: how can the self deceive itself, does that not require that it knows what it does not know? But the contradiction is easily resolved by defining the self as the conscious mind, so that self-deception occurs when the self is being kept in the dark, when the larger organism preferentially keeps true information out of consciousness, and when it misleads the conscious mind. Sometimes, this may involve activities of the conscious mind itself, which arrange - e.g., via active memory suppression - for later mental states to be biased in a particular way but usually the processes are unconscious and bias conscious mentation in a great variety of ways (see Gilbert 2006). So the key to defining self-deception is that true information is preferentially excluded from consciousness and, if held at all, is held in (varying degrees of) unconsciousness.

A very dramatic example of self-deception was demonstrated experimentally almost 30 years ago, with true and false information simultaneously stored within individuals but showing a strong bias toward the true information hidden in the unconscious mind. Gur and Sackeim (1979) also showed that they could manipulate an individual’s self-deception - so defined - by the simple device of increasing or decreasing the individuals’ opinion of themselves.

The experiment was based on a simple fact of human biology: we are physiologically aroused by the sound of the human voice but more so to the sound of our own voice (as reproduced, for example, on a tape recorder). We are unconscious of these effects. Thus, you can play a game of self-recognition, in which people are asked whether a voice is their own or not (conscious self-recognition) at the same time recording whether unconscious self-recognition was achieved, via higher arousal.

Here is how it worked. People were matched for age and sex and asked to read the same paragraph from Thomas Kuhn’s “Structure of Scientific Revolutions,” then these recordings were chopped into 2, 4, 6, 12, and 24 s segments, and a master tape was created consisting of a mixture of own and other voices. Meantime, the individual is hooked up to a machine measuring his or her galvanic skin response (GSR), which is normally twice as great for hearing one’s own voice as hearing someone else’s. People are asked to press a button to indicate if they think the recording is of self and another button to indicate how sure they are.

Several interesting things were discovered. First, some people denied their own voices some of the time, this was the only kind of mistake they made, and they seemed to be unconscious of making it (when interviewed later, only one was aware of having made this kind of mistake), yet the skin had it correct, that is, showed the large increase in GSR expected upon hearing one’s own voice. By contrast, another set of people heard themselves talking when they were not - they projected their voice - and half were aware later that they had sometimes made this kind of mistake, but the skin once again had it correct. There were two other categories, those who never made mistakes and those who made both kinds, sometimes even fooling their skin; but in what follows, we neglected these two categories.

This is unconscious self-recognition shown to be superior to conscious recognition, but Gur and Sackeim also showed that one could affect the tendency toward self-denial or projection by manipulating the person’s opinion of himself/herself. Made to feel bad by a poor score (in fact, randomly administered) on a pseudo-exam, individuals started to deny their voices more often. Made to feel good by a good score, individuals started to project their voices. It was as if self-presentation was expanding under success and contracting in response to failure.

Another interesting feature - never analyzed statistically - was that deniers also showed the highest GSRs to all stimuli. It was as if they were primed to respond quickly, to deny the reality and get it out of sight, while inventing the reality (projecting) seemed a more relaxed enterprise. Perhaps, reality that needs to be denied is more threatening than is the absence of reality one wishes to construct - or, in any case, denial can be dealt with more quickly.

Although this example is nice in showing clear, strong effects, it is not the only paradigm for self-deception. It is not necessary, for example, for the true information to be simultaneously stored. We act very early in information-generating processes to produce biased results. For example, we will choose not to read factual articles on a position we oppose (e.g., that marijuana consumption has negative personal effects) but to read those consistent with our views (e.g., marijuana consumption is beneficial). Somewhere in our body may be stored facts regarding preferential attention but not the content of the missed articles.

Alas! except for a couple of trivial articles quibbling with aspects of Gur and Sackeim’s work (e.g., Douglas and Gibbons 1983), no follow-up work has
appeared. How do those who make no mistakes differ from those who deny or project? And why do some individuals make both kinds of errors? Are deniers really more aroused, in general, than the other categories of people? And which voices of one’s own do we deny and which expropriate from others, and why? Unfortunately, we have no answers to these or a host of related questions. In part, this reflects the difficulty of the work itself, very painstaking and demanding, especially, in the pre-computer era, but mostly it reflects the degree to which self-deception has not been seen as a major subject in psychology, for which this was a significant methodological breakthrough. In any case, it is a shame: Gur and Sackeim created a whole new line of work which then failed to develop. The fact that is has failed to develop is cited as evidence against the original work. One prominent social psychologist told me that the results were so uncertain in his field that unless there was at least one replication, no one paid attention to solitary results.

18.6 One Needs a Separate Theory of Reality

Since deceit and self-deception must always be judged against the truth, one needs a separate theory of reality that is reliable against which to test self-deception. Evolutionary theory, of course, pretends to provide exactly that. The dangers of a false theory of reality are illustrated by the failure of both Freud and Marx’s theories of self-deception. Although Freud was able to describe such phenomena as denial, repression, and projection, his own unfounded theory of human development led him to deny a thing and project the other. For example, he denied that sexual advances from male relatives or step-relatives were a common problem for females, and projected on these women, the desire for exactly such encounters. An evolutionary approach is not congenial to the notion that women should have an inborn desire for sexual congress with close male relatives. Quite the contrary, but with a critical asymmetry, males are more likely to benefit genetically from such inbreeding than are the investing females. Here Freud’s absence of any plausible view of human development allowed him to twist his self-deception argument any which way. Marx, in turn, provided an analysis of bourgeois deceit and self-deception but his naïve theory of inevitable economic evolution only encouraged socialist self-deception.

18.7 Is Self-Deception the Psyche's Immune System?

The immune analogy is very popular within psychology. The argument goes as follows. Just as our body is under constant threat from parasites, so is our psyche under threat from factors that reduce happiness. Hence, we have psychological defense mechanisms, just as we have immunological ones, the one to keep us healthy and disease-free, the other to keep us happy. In one formulation, people are seen as having a “psychological immune system that defends the mind against unhappiness in much the same way that the physical immune system defends the body against illness” (Gilbert 2006).

This is said to be an “unusually appropriate” analogy because immune response and the degree of psychological defensiveness both share the trait that too little is bad but so is too much. Alas, this is true of all biological systems. Otherwise, we would not have stable phenotypes. Too little oxygen is bad, and so is too much. Too little food is bad – as is too much. And right down the line, height, weight, salt, water, curvature of your left thumb, tendency to visit dentists, everything is bad in its extremes.

What then is the image of how this immune-like defensive system works? “We need to be defended – not defenseless or defensive – and thus our minds naturally look for the best view of things while simultaneously insisting that those views stick reasonably closely to the facts” (Gilbert 2006).

“Reasonability” is the operative word here, undefined and as elastic as you could want. We are seen as keeping ourselves happy in good part via self-deception: denial, projection, disassociation, and so on. We cook the facts, we bias the logic, we overlook the alternatives – in short, we lie to ourselves. Meanwhile, we apparently have a “reasonability center” that, by unknown criteria, determines just how far we will be permitted to protect our happiness via self-deception (without, for example, looking ridiculous to others). Why was evolution unable to produce a more sensible way of regulating such an important emotion as happiness? Something more reasonable?

Contrast the real immune system. It deals with a major problem common to all of life, that of parasites, organisms that eat us out from the inside. Parasites are often the major selection pressure every generation on their hosts, much stronger than that of predators, for example. The immune system uses a variety of direct reality-based molecular mechanisms to attack, disable, engulf, and kill a veritable zoo of invading organisms – thousands of species of viruses, bacteria, fungi, protozoa, and worms – themselves using techniques honed over hundreds of millions years of intense natural selection. The immune system also stores away an accurate and large library of previous attacks, with the appropriate counter-response now pre-programmed in advance.

The vertebrate immune system is at least 300 million years old and is extremely sophisticated (similar evidence is now emerging for the insect immune system). Several dozen cell types are produced and marshaled in a bewildering array of patterns by a wide range of neurochemicals. The system is also very costly, perhaps on the order of the brain itself. It generates several grapefruit volumes of new tissue every 2 weeks. When unnecessarily aroused through false antigens (e.g., sheep red blood cells), immune arousal compromises survival in nature, sometimes strongly. Because it is so costly, it also acts as a great reservoir of energy against which the larger system can borrow in times of need; for example, effects of stress and sex steroids act to depress immune function, presumably the energy is spent on more immediately pressing problems (whatever is causing the stress or sexual and aggressive opportunities).
What on earth does all of this have to do with the ego's need to defend itself against various "threats"? And what exactly are these a threat to? One's self-opinion? Why is that a matter of pressing survival value? Parasites threaten your life. And why adopt something as dubious as self-deception to solve this problem? Put another way, where does our immune system improve its function by lying to itself - why is there no analogy on this key point? Granted the selection pressures associated with social interactions have steadily increased in our lineage - perhaps, exploding with language - but self-image, self-esteem, and ego strength are surely a small part of all of this.

It is possible that psychology has gotten all of this terribly wrong for so long by simply taking seriously an inside-outside approach to life - introspection will show us the way - in which we, in effect, choose our self-deceptions as building blocks in our theory? Social psychology has wedded itself to a thoroughly defensive view of self-deception, one that is itself congenial to an inflated self-perception: I am not lying to myself the better to deceive you but rather I lie to myself to defend myself from attacks on my personal integrity, my very happiness.

18.8 Self-Deception Helps Fool Others, While Reducing the Cognitive Cost of Doing So

Imagine two animals squaring off in a physical conflict. Each is assessing its opponent's self-confidence along with its own - variables expected to predict the outcome some of the time. It is easy to imagine that biased information flow within the individual can facilitate false self-confidence, which some of the time will pay for itself by fooling the opponent. Nonverbal self-deception can be selected in aggressive and competitive situations, the better to fool antagonists. Much the same could be said for male/female courtship relations. A male's false self-confidence may give him a boost some of the time. A biased mental representation can be produced, by assumption, without language.

The above is meant to demonstrate that in at least two contexts - aggressive conflict and courtship - selection for deception may easily have favored self-deception; that is, biased information flow within an organism to its consciousness, even when no language is involved. There are undoubtedly many other such contexts, for example, parent/offspring. But the simple fact is that language must have greatly expanded the opportunities for deceit and self-deception. If one great virtue of language is its ability to make true statements about events distant in space and time, then surely one of its social drawbacks is its ability to make similarly distant false statements, so less easily contradicted than statements about the immediate world. Once you have language, you have an explicit theory of self, social relationships and the world, ready to communicate to others. Numbers of new true assertions possible are matched by an infinitely greater number of false assertions, and so on.

18.9 Four Examples of Self-Deception

Although we have few examples of self-deception with the simple clarity and power of Gur and Sackeim (1979), there is, in fact, an enormous literature on the subject, crossing several disciplines including social psychology, behavioral economics, animal behavior, neurophysiology, immunology, and the study of everyday life. I have chosen four examples. First is self-inflation, often measured verbally but now with new techniques at a deeper level. It appears to be very general in human life. Second is the neurophysiology of thought suppression, which combines spatial brain coordinates of ongoing mental activity (fMRI) with measures of success at suppressing material from consciousness, an activity known also to have immediate immune costs (Pennebaker 1997). Third is old-age positivity,
a series of cognitive biases that creep in by age 60 so as to give a rosier view of the social world than is warranted. Does this produce immune benefits, especially useful in later life? Finally, the study of monkey and ape behavior in the wild suggests that deceptive behavior is more common per unit time in species with relatively large neocortices (social brains). If this is true within species, we might expect brighter people to be, on average, more self-deceived, a possibility with some serious social implications.

18.10 Self-Inflation is the Rule in Life

Animal self-inflation routinely occurs in aggressive situations (size, confidence, color) as well as during courtship (same variables). Self-inflation is the dominant style in human psychological life (Greenwald 1980; Gilbert 2006), adaptive self-diminution appearing as an occasional variant (Hartung 1988). People routinely put themselves in the top half of positive distributions and the lower half of negative ones. Eighty percent of U.S. high school students place themselves in the top half of students in leadership ability, but, of course, for extreme examples of self-deception you can hardly beat academics: 94% of them in one survey place themselves in the top half of their profession. I plead guilty. Even when tied to a bed in some back mental ward of a hospital, I still believe I am performing better than half of my colleagues — and this is not only a comment on Rutgers University.

Subtler linguistic features of self-deception have been described. When describing a positive group effect, we adopt an active voice but when the effect is negative, we unconsciously shift to a passive voice: this happened and then that happened and then costs rained down on all of us (reviewed in Greenwald 1980). When in-group members do something positive, we tend to make a general statement, “she is a good person,” just as we do when out-group members do something negative “he is wicked” but when an in-group member does something bad, we tend to describe it precisely, “she stepped on my toes,” just as we do when an out-group member does something good, “she gave me directions to the Bahnhof” (Maass 1999).

A recent methodology permits a very striking result (Epley and Whitchurch 2008). With the help of a computer, individual photos were morphed either 20% toward attractive faces or 20% toward unattractive. Among other striking effects, when the individual tries quickly to locate his or her real face, its 20% positive or its 20% negative in a background sea of 11 faces of other people, people were quickest to spot the positive face (1.86 s), less so for the real face (2.08 s), and slowest for the ugly one (2.16 s). On average, increasing degree of attractiveness improved the speed of perception by 1/10th of a second out of 2 seconds. The beauty of this is that there has not been the usual verbal filter — what do you think of yourself? — but only a measure of speed of perception.

18.11 The Neurophysiology of Thought Suppression

One particular kind of self-deception – consciously mediated efforts at suppressing true information from consciousness – has been studied neurophysiologically in a most revealing way (Anderson et al. 2004). Different sections of the brain may have been coopted in evolution to suppress the activity of other sections in order to create self-deceptive thinking.

Consider, for example, the active conscious suppression of memory. At times, we actively attempt to suppress thoughts. In the laboratory, individuals are instructed to forget an arbitrary set of symbols that they have just learned. The effect of such efforts is highly variable, measured as the degree of forgetting achieved a month later and this variation is associated with variation in the underlying neurophysiology. The more highly the dorsolateral prefrontal area is activated during directed forgetting, the more it suppresses the ongoing activity in the hippocampus (where memories are typically stored) and the less is remembered a month later. The dorsolateral prefrontal area is otherwise often involved in overcoming cognitive obstacles and in prepotent motor activity, that is, preparing for physical activity (muscle movement). It is tempting to speculate that this area of the brain was coopted for the new function of suppressing memories because it was often involved in affecting other brain areas, i.e., activating behavior. It may be unrelated but whenever I experience an unwanted thought and act at once to suppress it, I often experience an involuntary twitch in my arms, as if trying to push something down (and out of sight).

18.12 Old-Age Positivity and Immune Function

There is a striking bias toward positive social memories and perceptions that sets in by age 60 (and perhaps somewhat earlier). At ages 20–30, the human shows no tendency to remember faces with positive expressions more often than those with negative ones, or to spend more time examining such pictures. But by age 60, a bias is apparent: positive faces are remembered more readily and they are attended to more carefully (Mather and Carstensen 2005). When a dot is presented on the side of a screen at which a positive face is presented, the dot is perceived more quickly if it succeeds the positive face (and less slowly if it succeeds a negative one, compared with neutral Mather and Carstensen 2003). This involves a measurable effect in the amygdala, where positive faces evoke a stronger response than negative in older people but not in younger people (Mather et al. 2004). Older people, compared to younger, are more likely to respond to a musically induced negative mood by preferentially looking at positive faces, as if actively inducing a positive mood (Isaakowitz et al. 2008).

Why show such a positivity bias? Half the problem is trivial. Young people will be wise to pay attention to reality, both positive and negative, the better to make the
appropriate responses later; by old age, it hardly matters what you learn and since
greater positive affect is associated with stronger immune response (Rosenkranz
et al. 2003), you may trade grasp of reality for a boost in dealing with your main
problem, that of parasites and cancer. A positivity bias sacrifices learning in the
future concerning negative outcomes the better to enjoy strong immune function
now. Grandchildren may admire gramps and grandma because nothing seems to
faze them, but gramps and grandma might be living in a positivity-enveloped world,
the better to deal with their internal enemies.

It is an interesting coincidence that although our implicit bias in favor of youth
over old age hardly changes with age – from 20 to 70, we favor young over old – by
our 40s, our explicit bias in favor of youth declines until at exactly 60 we start to
prefer older to younger – like everyone else we implicitly associate youth with
positive features, but we start preaching the opposite at the same time at which we
ourselves display the old-age positivity bias (Nosek et al. 2002).

18.13 Are Intelligence and Self-Deception Correlated?

It is easy to imagine that intelligence and consciousness are two independent axes
of human behavior, perhaps equally important, but uncorrelated. Thus, one can be
bright and deluded or slow and honest with all combinations equally likely.
Likewise, it is easy to imagine that the two axes are positively correlated. The
smarter you are the less self-deception (greater consciousness). Your innate superi­
ority in intellectual power can easily be turned back on your deceptive tendencies,
so that you see through your lies and adjust appropriately. But what does that
mean – you lie less or lie more?

I do not believe that degree of self-deception and intelligence are uncorrelated or
that they are negatively correlated. I believe quite the opposite. Degree of con­
sciousness and intelligence are positively correlated: brighter people are more
likely to act deceptively and to practice self-deception. This increases the chance
that the net effect of their actions will be negative instead of positive. This is, to put
it mildly, an underemphasized underbelly of high intelligence. Of course, there are
exceptions. It is not surprising that the academically less gifted are more likely
to cheat (and thus act deceitfully) as indeed they are.

One line of evidence comes from monkeys and apes. The size of the neocortex –
or better still, its relative portion of total brain size – is positively associated with
use of tactical deception in nature, where tactical deception includes any kind of
deception that can be seen to give an advantage. A large list of appropriate acts from
nature was assembled from the primary scientific literature and used to solicit a still
larger sample from active scientists. Study effort and group size were controlled as
were taxonomic effects. Conclusion: since neocortex size is correlated with intelli­
gence – including social intelligence – across a broad range of monkeys and apes,
we know that deception occurs more often the smarter the species is. So, perhaps,
does self-deception.

Another line of evidence comes from children. As children mature, they become
increasingly intelligent and increasingly deceptive. This is not an accident. The
very maturing capacity that gives them greater general intelligence also gives them
greater ability to suppress their behavior and create a novel behavior. There is also a
clear evidence that natural variation in intelligence, corrected for age, is positively
associated with deception in children (Lewis unpubl. data) using the peek/no peek­
lie/not lie paradigm that has been used to such good effect (e.g., Crossman and
Lewis 2006). Even health of the child at birth (as measured by a weighted sum of 32
perinatal factors) is positively correlated with lying.

Thus, if you wish to cherish a self-image that you are smarter than average or
even that your group is, you may also need to imagine that you (and the group) are
more prone to deceit and self-deception, with net effect on others uncertain.

18.14 Imposed Self-Deception

So far we have spoken of self-deception evolving in the service of the actor. This is
the natural first step, but we are also highly sensitive to others, to their opinions,
desires, actions, and so on. More to the point, we can be manipulated and dominated
by them. This can result in self-deception being imposed on us by others (with
varying degrees of force). Extreme examples are instructive. A captive may come
to identify with his captor, an abused wife may take on the world-view of her
abuser, molested children may blame themselves for the transgressions against
them and the resulting misery. These are cases of imposed self-deception and if
they are acting functionally from the standpoint of the victimized (by no means
certain), then they probably do so by reducing conflict with the dominant individual.
At least this is often the theory of the participants themselves: an abused wife may
be deeply frightened and rationalizes acquiescence as the path least likely to
provocate additional severe assaults, this is soon most effective if actually believed.

Let us consider another example of imposed self-deception, one with deeper
social implications. It is possible to measure something called a person’s “explicit”
self-preference as well as an “implicit” one. The explicit simply asks people to state
their preferences directly, e.g., for the so-called “black” people over “white” (to use
the degraded language of the United States) where the actor is one or the other. The
implicit measure is more subtle. It asks people to push a right-hand button for black
or “good” stimuli (e.g., positive words) and left for white or bad ones – and then
reverses everything, black or bad, white or good. We now look at latencies – how
long does it take an individual to respond when they must punch white or bad versus
white or good – and assume that shorter latencies (quicker responses) means the
terms are, by implication, more strongly associated in the brain. Hence, the term
“implicit association test” (IAT), invented only in 1998 (Greenwald et al.), has now
generated an enormous literature, including (unusual for the social sciences) actual
improvements in methodology (Greenwald et al. 2003). There are several websites
that harvest enormous volumes of data over the internet (e.g., at Harvard, Yale, and
the University of Washington), and these studies have produced some striking findings (Nosek et al. 2002).

For example, black and white people were similar in their explicit tendency to value self over other, blacks if anything more strongly so, but when it came to the implicit measures, whites were even more strongly in their own favor than they were explicitly, while blacks—on average—preferred white over black, not by a huge margin but, nevertheless, they preferred other to self (Nosek et al. 2002). This is most unexpected according to evolutionary theory, where self is the beginning (if not end) of self-interest.

This has the earmarks of an imposed self-deception—valuing yourself less than you do others—and it may come with some negative consequences. For example, priming black students for their ethnicity strongly impairs performance. This was indeed one of the first of what are now hundreds of demonstrations of this “priming” effect. Black and white undergraduates at Stanford arrived in a lab to take relatively difficult aptitude tests. In one situation, the students were simply given the exams; in the other, each was asked to give a few personal facts, one of which was their own ethnicity. Black and white scored equally well with no prime. With a prime, whites did slightly (but not significantly) better while blacks’ scores plummeted by nearly 1/2. You can even manipulate one person’s performance in opposite directions by giving opposing primes: Asian women perform better on math tests when primed with “Asian” and worse when primed with “woman.” No one knows how long the effect of such primes endures but nor does anyone know how often a prime appears: how often is an African-American reminded that he or she is such? Once a day? Every half hour? Once a month? I think the number is somewhere between the second and the third.

The strong suggestions then is that it is possible for a historically degraded and/or despised minority group, now subordinate, to have an implicit self-image that is negative, to prefer other to self—indeed, oppressor to self—and to underperform as soon as made conscious of the subordinate identity. This suggests the power of imposed or induced self-deception—some or, indeed, many subordinate individuals adopting the dominant stereotype regarding themselves. Not all of course, and the latter presumably more likely to oppose their subjugation since they are conscious of it. In any case, revolutionary moments often seem to occur in history when large numbers of individuals have a change in consciousness—regarding themselves and their status. Whether there is an accompanying change in IAT is another matter.

One more form of induced self-deception is worth mentioning. It is surprisingly easy to convince people to make false confessions to major crimes even though this may—and surprisingly often does—result in incarceration for long periods of time. All that is required is a susceptible victim and good old-fashioned police work applied 24/7: isolation of the victim from others, sleep deprivation, coercive interrogation in which denial and refutation are not permitted, false facts provided and hypothetical stories told (we have your blood on the murder weapon, perhaps you woke in a state of semi-consciousness and killed your parents without intending or being aware of it etc., with the implication that a confession will end their interrogation). People vary in in the susceptibility range they are to these pressures and in how much self-deception is eventually induced. Some go on to create false memories to back up their false confessions (Kassin 2008).

There is also a kind of self-deception that could be called imposed self-deception, but which could also be considered defensive self-deception. Consider an individual being tortured. The pain can be so great that something called “disassociation” may occur: the pain is separated off from other mental systems, presumably so as to reduce its intensity. As if the psyche or nervous system protects itself from severe pain by objectifying it, distancing it, and splitting it off from the rest of the system. One can think of this as being imposed by the torturer but also as a defensive reaction permitting immediate survival under most unfavorable circumstances. We know from many personal accounts that this is but a temporary solution and that the torture itself and utter helplessness against it endure as psychological and biological costs. There are, of course, more modest forms of “disassociation” from pain than that of torture, such as a mother distracting her child by tickling it.

18.15 Deceit and Self-Deception Seen as an Evolutionary Game

To model the evolution of deceit and self-deception in humans more exactly, I recommend pursuing the strategy that has proven so successful with reciprocal altruism and cooperation. That is, model opposing strategies as simple rules with specified costs and benefits in interaction with each other. This was first applied successfully to the evolution of reciprocal altruism by Axelrod and Hamilton (1981) modeled as a successive series of simultaneous moves with only two options for each of the two players, cooperate or defect, each combination with specified pay-offs (the so-called iterated Prisoner’s Dilemma).

It was shown that a very simple strategy—tit-for-tat—beat out all others in computer tournaments and is a fairly robust strategy in evolutionary games: be cooperative on the first move, and imitate your partner’s previous move on your next one. In short, reward cooperation with cooperation and punish defection with defection. An advantage of casting the problem in terms of games is that these can be played both mathematically and for real. When their patterns are convergent as they are here, we can have even greater confidence in the underlying logic.

Once the simple tit-for-tat strategy was described, certain problems were discovered that required modified strategies (briefly reviewed in Trivers 2005). For example, occasional errors can lock two tit-for-tatters into a most unfortunate situation, endless “vendettas” in which each reverses its move exactly out of synchrony with the other, never achieving simultaneous cooperation, thereby greatly reducing the success of the tit-for-tat strategy. “Generous tit-for-tat” solved this problem by allowing a tit-for-tatter occasionally (say 1/3rd of the time) to cooperate after a defection by the partner—small cost, larger gain under a variety of realistic conditions (Nowak and Sigmund 2004). Later, “win-stay, lose-shift” was shown to be superior still against a background of tit-for-tatters, defectors, and
generous tit-for-tatters (Nowak and Sigmund 1993). Observer effects can also be modeled (Nowak and Sigmund 1998a,b), leading to "indirect reciprocity." In this case, my strategy towards you depends not only on what you have done to me but also on what you have been observed doing to others.

The simplest application of the above to deceit would be to treat it as a Prisoner's Dilemma. Two individuals can each tell the other the truth (both cooperate) or lie (both defect) or one of each. There are two problems with this. One is that a critical new variable becomes important: who believes whom? If I believe you and you are lying, I suffer. If you lie and I disbelieve you, you suffer. By contrast, in the Prisoner's Dilemma, each individual knows after each reciprocal play how the other played (cooperate or defect) and tit-for-tat and its elaborations provide a simple reciprocal mechanism that can operate under the humblest of conditions—in bacteria. The second problem is that with deception, there is no obvious reciprocal logic. If you lie to me, this does not mean my best strategy is to lie back to you—it usually means that my best strategy is to distance myself from you or punish you.

Karl Sigmund (pers. comm.) has suggested that it might be useful to adapt the Ultimatum Game to this problem. In the UG, a proposer offers a given split of (say) $100: e.g., $80 to self, $20 to the responder. The responder, in turn, can accept the split, in which case the money is split accordingly or the responder can reject the offer, in which case neither receives anything. Often the game is played as a one-shot anonymous encounter, i.e., individuals play only once with whom they do not know and with whom they will not interact in the future.

Sigmund argues as follows. Imagine a modified UG in which there are two possible pots (say $100 and $400) and both players know this. One pot is then randomly assigned to the proposer. Now let us say the proposer offers you $40; this could represent 40% of the pot (in which case you should accept) or 10% (most people would reject). The proposer is permitted to lie and tell you that the pot is the smaller of the two when, in fact, it is the larger. You can trust the proposer not but the key is that you are permitted to pay to find out the truth from a (disinterested) third party. (This measures the value you place in reducing your uncertainty regarding the proposer's honesty). If you then discover that the proposer lied, you should have a moral (or, at least, moralistic) motive to reject the offer, and the other way around, for the truth (all compared to uncertainty, i.e., not paying to find out). Note that from a purely economic point of view, there is no benefit in finding out the truth, since it costs money after which it may lead to an (otherwise) unnecessary loss of whatever is offered. In Sigmund's words: "how much would a responder be willing to pay for reducing the uncertainty and go for a possibly inconvenient third party?" (This measures the value you place in reducing your uncertainty regarding the proposer’s honesty). If you then discover that the proposer lied, you should have a moral (or, at least, moralistic) motive to reject the offer, and the other way around, for the truth (all compared to uncertainty, i.e., not paying to find out).

Note that from a purely economic point of view, there is no benefit in finding out the truth, since it costs money after which it may lead to an (otherwise) unnecessary loss of whatever is offered. In Sigmund's words: "how much would a responder be willing to pay for reducing the uncertainty and go for a possibly inconvenient third party?" Note that the game can be played in real life with varying degrees of anonymity and also multiple times, as in the iterated Prisoner's Dilemma. As ability to discriminate develops, the other person will benefit more from your honesty (quickly seen as such) and suffer less from deception (spotted and discarded). When people are in greater need, they may be expected not to pay to find out the truth but rather to accept the offer whatever its relative size is.

When we add self-deception, a possible game quickly becomes very complicated. One can imagine actors who are stone-cold honest (cost: information given

away, naive regarding deception by others), consciously dishonest to a high degree but with low self-deception (cost: higher cognitive cost and when detected), dishonest with high self-deception (more superficially convincing at lower immediate cognitive cost but suffering later defects and acting more often in the service of others, and so on). Without anything else to offer along these lines, I suggest that those talented at the mathematics of simple games or studying them via computer simulations might find it rewarding to define a set of characters along the lines just mentioned, and then assign variable quantitative effects so as to explore their combined evolutionary trajectory. Perhaps, results will be trivial and trajectories will depend completely on the relative quantitative effects assigned but more likely deeper connections will emerge, seen only when the coevolutionary struggle is formulated explicitly.

18.16 The Cost of Deconstructing Lies

The cost of seeing through deception is not trivial, and in some cases, it is substantial. Think of the daily drumbeat of propaganda emerging from the government in times of war or class warfare (2001–2006 in the U.S., for example). On a personal level, I was first exposed to this when I was about 6 years old. Having saved 6 dollars over the space of 2 months to buy a knife displayed in a window, I showed up, only to be told that I was 1 dollar short. Nonsense, I said, "the sign outside says "$6." The shopkeeper took me to the sign and showed me that written after the 6 in very small letters was 98, i.e., almost 1 dollar. I was incredulous—and very angry—how did it make sense, I wanted to know, for him to misrepresent the true value of an item by subtracting two pennies so as to generate just the kind of mistake I had, in fact, made? He assured me that the practice was widespread. I soon confirmed that this was true. Almost all prices—gas, food, furniture—when spotted at a distance, appeared to be one full unit below what they actually were. For a couple of weeks I walked around in a daze, benumbed at the amount of unnecessary arithmetical calculation this system required: always adding a unit or two to the total in order to calculate the real value. How was it possible, I kept asking myself, that this was the system of posting prices that had developed?

And, I think, there was my mistake. This was not a rational system agreed upon by all actors—or what Jesus might have told us to do—this was the system that had actually developed over time. In a nutshell, honesty is not evolutionarily stable. It is easily displaced by deception which, in turn, forms a new equilibrium. Further, deception can be counter-selected but so may be a return to honesty, since honest valuations will often be devalued along with those that are hyped. That is, we will frequently add a unit to honest prices, decline to buy them, and the honest shopkeeper suffers.

So also with self-deception. Over evolutionary time, we have been driven downhill by selection such that a degree of self-deception is common to most or all of us. It has formed a new equilibrium such that honest people may come off as
lacking, in part because others unconsciously compensate in their estimates for the expected degree of inflation. No inflation so we make them smaller than they really are. And, perhaps, the happiness thermostat in our body has been reset so as to assume a certain degree of self-deception.

References


Hauser MD (1992) Costs of deception: cheaters are punished in rhesus monkeys (Macaca mulatta). Proc Natl Acad Sci USA 89:12137-12139


