

The Elements of a Scientific Theory of Self-Deception

ROBERT TRIVERS

*Department of Anthropology, Rutgers University, 131 George Street,
New Brunswick, New Jersey 08901-1414, USA*

ABSTRACT: An evolutionary theory of self-deception—the active misrepresentation of reality to the conscious mind—suggests that there may be multiple sources of self-deception in our own species, with important interactions between them. Self-deception (along with internal conflict and fragmentation) may serve to improve deception of others; this may include denial of ongoing deception, self-inflation, ego-biased social theory, false narratives of intention, and a conscious mind that operates via denial and projection to create a self-serving world. Self-deception may also result from internal representations of the voices of significant others, including parents, and may come from internal genetic conflict, the most important for our species arising from differentially imprinted maternal and paternal genes. Selection also favors suppressing negative phenotypic traits. Finally, a positive form of self-deception may serve to orient the organism favorably toward the future. Self-deception can be analyzed in groups and is done so here with special attention to its costs.

INTRODUCTION

An important component of a mature system of social theory is a sub-theory concerning self-deception (lying to oneself, or biased information flow within an individual, analogous to deception between individuals). This sub-theory can always be turned back on the main theory itself. There can be little doubt about the need for such a theory where our own species is concerned—and of the need for solid, scientific facts which bear on the theory. Whether through a study of one's own behavior and mentation (e.g., for a novelist's treatment¹) or of societal disasters (e.g., in aviation^{2,3} or misguided wars^{4,5}), or a review of findings from psychology,^{6–13} we know that processes of self-deception—active misrepresentation of reality to the conscious mind—are an everyday human occurrence, that struggling with one's own tendencies toward self-deception is usually a life-long enterprise, and that at the level of societies (as well as individuals) such tendencies can help produce major disasters (e.g., the U.S. war on Viet Nam). With potential costs so great, the question naturally arises: what evolutionary forces *favor* mechanisms of self-deception?

A theory of self-deception based on evolutionary biology requires that we explain how forces of natural selection working on individuals—and the genes within them—may have favored individual (and group) self-deception, where natural selection is understood to favor high inclusive fitness, roughly speaking, an individual's (or gene's) reproductive success (RS = number of surviving offspring) plus effects on the RS of relatives, devalued by the degrees of relatedness between actor and relatives.¹⁴ There is ample evidence that this simple principle provides a firm founda-

tion for a general theory of social interactions.¹⁵ Deception between individuals who are imperfectly related may often be favored when this gives an advantage in RS to the deceiver (see Refs. 15 and 16 for some examples) but the argument for *self*-deception is not so obvious.

For a solitary organism, the prospects seem difficult, if not hopeless. In trying to deal effectively with a complex, changing world, where is the benefit in misrepresenting reality to oneself? Only in interactions with other organisms, especially conspecifics, would several benefits seem to arise. Because deception is easily selected between individuals, it may also generate *self*-deception, the better to hide ongoing deception from detection by others.^{2,15,17} In this view, the conscious mind is, in part, a social front, maintained to deceive others—who more readily attend to its manifestations than to those of the actor's unconscious mind. At the same time, social processes, such as parent-offspring conflict¹⁸ in a species with a long period of juvenile dependency—or, a more general group-individual conflict—may generate conflicting internal voices, representing parental and own self-interests (or group and self), with consequent reality-distortion within the individual.¹⁸ For example, the parental view may be overstated internally (for example, via parental manipulation), requiring careful devaluation or counter-assertion.

A stronger force may arise from the fact that different sections of our genome (mtDNA, sex chromosomes, autosomes and, separately, the maternal and paternal chromosomes) often enjoy differing degrees of relatedness to others, with consequent internal conflict between the sections potentially generating deception within the individual, a kind of “selves-deception.”¹⁹ Internal conflict may occur for other reasons, as well, and may or may not involve biased information flow. For example, it is certain that all of us possess disadvantageous traits, both genetic and developmental, and, thus, natural selection may have favored super-ordinate mechanisms for spotting negative traits in the phenotype (perhaps especially behavioral ones) and then attempting to suppress them. This may be experienced sometimes as internal psychological conflict and may or may not involve biased information flow. Finally, a positive stance toward life may have intrinsic benefits (and not only for social species). A concentration on the future—and positive outcomes therein—may benefit from seeing past setbacks as blessings in disguise and the current path chosen as the best available option. In short, positive illusions may give intrinsic benefit.^{8,9} Is this self-deception or merely optimism in the service of reproductive success?²¹

SELF-DECEPTION IN THE SERVICE OF DECEIT

One model for internal fragmentation and conflict is represented in FIGURE 1. True and false information is simultaneously stored in an organism with a bias towards the true information's being stored in the unconscious mind, the false in the conscious. And, it is argued, this way of organizing knowledge is oriented towards an outside observer, who sees first the conscious mind and its productions and only later spots true information hidden in the other's unconscious. This is self-deception in the service of deception of others. It may be expected to flourish in at least the following five kinds of situations.

1. Denial of ongoing deception. Being unconscious of ongoing deception may more deeply hide the deception. Conscious deceivers will often be under the stress

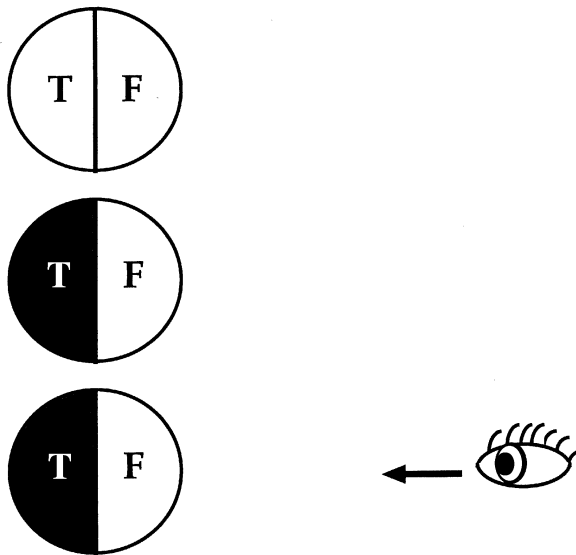


FIGURE 1. True (T) and false (F) information is simultaneously stored within an organism, but with a bias: the true is stored in the *unconscious* mind (shaded section), the better to deceive an on-looker (eye).

that accompanies attempted deception. Evidence from other animals suggests that, as in humans, deception, when detected, may often be met with hostile and aggressive actions by others.^{22–25} Thus, if I were in front of you now, lying to you about something you actually care about, you might pay attention to my eye movements, the quality of my voice, and the sweat on the palms of my hands (if you can reach them) as a means of detecting the stress accompanying deception, but if I am unconscious of the deception being perpetrated all these avenues will be unavailable to you.

2. *Unconscious modules involving deception.* In the above example, the main activity—verbal persuasion directed at others—is deceptive, but there are also situations in which your dominant activity (say, lecturing) is honest, but a minor activity is deceitful (stealing the chalk). These can be thought of as directed by unconscious modules favored by selection so as to allow us to pursue surreptitiously strategies we would wish to deny to others. Naturally these will often remain unconscious to us. I will shortly describe in detail a deceitful little module in my own life which I have discovered primarily because my pockets fill up with contraband: hard, concrete objects that others may soon miss. What is the chance that I perform numerous unconscious selfish modules whose social benefits do not pile up in one place, where I can notice them (and others confirm them), e.g., ploys of unconscious manipulation of others (including, of course, as an academic, expropriating their *ideas*)?

I have discovered over the years that I am an unconscious petty thief. I steal small, useful objects: pencils, pens, matches, lighters and other useful objects easy to pocket. I am completely unconscious of this activity while it is happening. I am, of course, now richly aware of it in retrospect, but after at least 40 years of performing

the behavior I am still unconscious ahead of time, during the action, and immediately afterwards. Perhaps because the trait is so unconscious, it appears now to have a life of its own and often seems to act directly against even my narrow interests. For example, I steal chalk from myself while lecturing and am left with no chalk with which to lecture (nor do I have a blackboard at home). I steal pencils and pens from my office and, in turn, from my home, so if I download my pockets at either destination, as I commonly do, I risk being without writing implements at the other end. Recently I stole the complete set of keys of a Jamaican school principal off of his desk between us. And so on.

In summary, noteworthy features of this module are that: (1) it is little changed over the course of my life; (2) increasing consciousness of the behavior *after* the behavior has done little or nothing to increase consciousness during or in advance of the behavior; and (3) the behavior seems increasingly to misfire, that is, to fail to steal *useful* objects.

What is the benefit of keeping this petty thievery unconscious? On the one hand, if challenged, I can act surprised and be confident in my assertion that nothing like this was ever my conscious intention (see below). On the other hand, unconsciousness ensures that my thievery will not interfere with ongoing behavior, while the piece of brain devoted to stealing can concentrate on the problem at hand, i.e., snatching the desired item undetected. Part of its consciousness has to be devoted to studying my *own* behavior since integrating its thievery into my other behavior will presumably make this harder to detect by others, including myself.

Incidentally, I believe I never, or almost never, pilfer from someone's office when it is empty. I have seen a choice pen and have seen my hand move toward it but I immediately stop myself and say, "but, Bob, that would be stealing," and I stop. Perhaps if I steal from you in front of your face I unconsciously imagine you have provided some acquiescence, if not actual approval. When I stole the principal's keys, I believe I was simultaneously handing him repayment of a small amount and wondering if I were slightly overpaying. Perhaps I reasoned to myself, "Well, this is for you, so *this* must be for me."

3. *Self-deception as self-promotion.* Another major source of self-deception has to do with self-promotion, self-exaggeration on the positive side, denial on the negative, all in the name of producing an image that we are "beneffective," to use Anthony Greenwald's⁷ apt term, toward others. That is, we benefit others and are effective when we do it. If you ask high school seniors in the United States to rank themselves on leadership ability, fully 80% say they have better than average abilities, but for true feats of self-deception you can hardly beat the academic profession. When you ask professors to rate themselves, an almost unanimous 94% say they are in the top half of the profession!²⁶ For many other examples, see Refs. 7 and 13. Tricks of the trade are biased memory, biased computation, changing from active to passive voice when changing from describing positive to negative outcomes, and so on.

4. *The construction of biased social theory.* We all have social theories. We have a theory of our marriages. Husband and wife, for example, may agree that one party is a long-suffering altruist, while the other is hopelessly selfish, but they may disagree over which is which.¹⁵ We each tend to have a theory regarding our employment. Are we an exploited worker, underpaid and underappreciated for value given (and fully justified in minimizing output and stealing company property)? We usually have a theory regarding our larger society as well. Are the wealthy unfairly in-

creasing their own resources at the expense of the rest of us? Does democracy permit us to reassert our power at regular intervals? Is the judicial system systematically biased against our kind of people (African-Americans for example)? The capacity for these kinds of theories presumably evolved in part to detect cheating in our relationships and in the larger system of reciprocal altruism.

Social theory is easily expected to be biased in favor of the speaker. Social theory inevitably embraces a complex array of facts and these may be very partially remembered and very poorly organized, the better to construct a consistent self-serving body of social theory. Contradictions may be far afield and difficult to detect. When Republicans in the House of Representatives bemoaned what the Founding Fathers would have thought had they known that a successor President was having sex with an intern, the Black American comedian Chris Rock replied that the Founding Fathers were not having intercourse with their interns, they were having intercourse with their *slaves*! This kind of undercuts the moral force of the argument given (for recent evidence supporting his assertion, see Ref. 27).

Alexander¹⁷ was, I think, the first person to point out that group selection thinking—the mistaken belief that natural selection favors things that are good for the group or the species—is just the kind of social theory you would expect to be promulgated in a group-living species whose members are concerned to increase each other's group orientation.

5. *Fictitious narratives of intention.* Just as we can misremember the past in a self-serving way, so we can be unconscious of ongoing motivation, instead experiencing a conscious stream of thoughts which may act, in part, as rationalizations for what we are doing, all of which is immediately available verbally should we be challenged by others: "But I wasn't thinking that at all, I was thinking such-and-such." A common form in myself is that I wish to go to point C, but can not justify the expense and time. I leap, however, at a chance to go to point B, which brings me close enough to point C so that, when there, I can easily justify the extra distance to C, but I do not think of C until I reach B. We may have much deeper patterns of motivation which may remain unconscious, or nearly so, for much longer periods of time, unconscious patterns of motivation in relationships, for example.

In summary, the hallmark of self-deception in the service of deceit is the denial of deception, the unconscious running of selfish and deceitful ploys, the creation of a public persona as an altruist and a person benefactive in the lives of others, the creation of self-serving social theories and biased internal narratives of ongoing behavior which hide true intention. The symptom is a biased system of information flow, with the conscious mind devoted, in part, to constructing a false image and at the same time being unaware of contravening behavior and evidence. The general cost of self-deception, then, is misapprehension of reality, especially social, and an inefficient, fragmented mental system. For a deeper view of these processes we must remember that the mind is not divided into conscious and unconscious, but into differing degrees of consciousness. We can deny reality and then deny the denial, and so on, *ad infinitum*. Consciousness comes in many, many degrees and forms. We can feel anxious and not know why. We can be aware that someone in a group means us no good, but not know who. We can know who, but not why, and so on.

The examples in this article are all taken from human life. While language greatly increases the possibilities for deceit and self-deception in our species, selection probably favored deception in social species for hundreds of millions of years and

TABLE 1. Consciousness (neuronal times)^a

			Time
Finger	⇒	Brain	20 ms
Round trip			50 ms
Sensation	⇒	Consciousness	500 ms
Round trip	+	Cognitive processing	100–200 ms
Neuronal start of act	⇒	Conscious “intention”	350 ms
“Intention”	⇒	Action	200 ms

this may have selected for some mechanisms of self-deception. Two animals evaluating each other in an aggressive encounter (or even in courtship) will be selected to pay close attention to the other individual’s apparent self-evaluation and level of motivation, both of which can be boosted by selective forgetting, as in humans.²⁸ In humans the major sex hormones (e.g., testosterone and estradiol) seem to be positively associated with degree of self-inflation.²⁸ Since testosterone is sometimes positively associated with aggression and aggression with self-deception (see below) such connection may make functional sense in both humans and other animals⁵ (where it could easily be pursued experimentally).

NEURONAL TIMES IN CONSCIOUSNESS

It is common to imagine that our conscious mind occupies a central place in our life, where apprehension of reality and subsequent decision-making is concerned. It is easy to imagine that information reaching our brain is immediately registered in consciousness and likewise that signals to initiate activity originate in the conscious mind. Of course, unconscious processes go on at the same time and unconscious processes may affect the conscious mind but there is not a great deal of time, for example, for something like denial to operate, certainly not if this requires spotting a signal and then, before it can reach consciousness, shunting it aside. And, voluntary activity, of which we are conscious as we act, may be affected by unconscious factors, but nevertheless plays the overriding role in directing activity. This is the conventional (pre-Freudian) view.

Thirty years of accumulating evidence from neurophysiology suggests that this is an illusion (TABLE 1). The first and, perhaps, most startling fact is that while it takes a nervous signal only about 20 ms to reach the brain, it requires a full 500 ms for a signal reaching the brain to register in consciousness! This is all the time in the world, so to speak, for emendations, changes, deletions, and enhancements to occur. Indeed, neurophysiologists have shown that stimuli, at least as late as 100 ms before an occurrence reaches consciousness, can affect the content of the experience.²⁹ Some additional times are the following.

It takes only 50 ms for a signal from the finger to cause, via a round-trip to the brain stem, the finger to be moved. Additional cognitive processing may require another 150 ms, but all of this is achieved without consciousness. Finally, what do we make of the following fact? 350 ms *before* we consciously intend to do something

the relevant neuronal activity begins and there is a further 200 ms delay after we “intend” to do something before we actually do it. It seems as if our conscious mind is more of an on-looker than a decision-maker.

THE LOGIC OF DENIAL AND PROJECTION

Denial and projection are basic psychological processes serving self-deception, though in slightly different ways. Sometimes we will wish to deny something, usually negative (e.g., that we have caused harm to others or “incriminating” personal facts regarding adultery, robbery or something shameful). At other times we may wish to project something onto others which is true of ourselves. In simple voice-recognition choice experiments (see below) denying one variable means choosing (or projecting) the other, and *vice versa*, but the two are distinguished by relevance to self (own voice more important than other). Projection and denial are likely to have different dynamics. Denial will easily engender denial of denial, the deeper to bury the falsehood. Denial may plausibly require a heightened level of arousal, the better to attend quickly to the facts needing denial and shunt them from consciousness. Projection, by contrast, may often be a more relaxed operation: it would be nice if the facts were true, but not critical if they are not.

These speculations are supported by the classic voice-recognition experiments of Gur and Sackeim,⁶ where unconscious self-recognition is measured by a relatively large jump in galvanic skin response (GSR). Some people deny their own voices some of the time, while others project their voice some of the time. In each case, the skin (GSR) has it right. Furthermore, when interviewed afterwards, almost all deniers deny their denial, while half of those projecting their voices are conscious after the fact that they sometimes made mistakes of exactly this sort. A comparison of the overall levels of GSRs shows that deniers exhibit the greatest GSRs to all stimuli, while projectors show the more relaxed profile typical of those who make no mistakes, as well as the hopelessly confused (those who both deny and project, sometimes fooling their own skin). Finally, Gur and Sackeim showed that denial and projection were motivated in a logical fashion: individuals made to feel bad about themselves started denying their own voices, while people made to feel better about themselves started projecting their own voices—as if self-presentation was being contracted or expanded according to relevant facts.

A student could go a long way by devising a series of follow-up experiments requiring only a tape recorder and a machine for reading the galvanic skin response. Is denial really associated with greater arousal than projection or correct apprehension of reality? What is happening with those individuals who make both kinds of mistakes—are they really completely confused some of the time? And if so, why? What kinds of voices of yourself do you deny after failure and which kinds do you project when you succeed (or believe you have)? If it is really true, as Douglas and Gibbins^{30,31} seem to show, that voices of familiar others evoke a GSR stronger than unfamiliar others, then what kinds of events cause us to deny familiar others? Is denial or projection more likely with increasing testosterone? And so on. Note that there is a large industry in the U.S. devoted to the use of lie-detector tests (which employ GSR as one if their measures), largely in the world of job interviews and job-related thefts. There is a parallel academic literature investigating the lie detector

TABLE 2. Homophobia scale: sample questions³²

1. I would feel comfortable working closely with a male homosexual
4. If a member of my sex made a sexual advance toward me, I would feel angry.
5. I would feel comfortable knowing that I was attractive to members of my own sex.
12. I would deny to members of my peer group that I had friends who were homosexual.
14. If I saw two men holding hands in public I would feel disgusted.
17. I would feel uncomfortable if I learned that my spouse or partner was attracted to members of his or her own sex.

methodology in various settings. It should be easy to integrate the study of self-deception into these studies. Indeed, it may be possible to see under what conditions self-deception decreases detection by others, both those using a lie-detector machine and those not. It would also be possible, in principle, to adapt their methodology to the study of self-deception in animals: birds, for example, may also show greater physiological arousal to their own or close relative's voice than to others and they could be trained to peck when they "thought" they heard their own voice instead of another's.²⁵

Denial of personal malfeasance may often strongly necessitate its projection onto others. Once years ago while driving, I took a corner too sharply and my one-year-old baby fell over in the back seat and started to cry. I heard myself harshly berating her nine-year-old sister (my stepdaughter)—as if she should know by now that I like to take my corners on two wheels, and brace the baby accordingly. But the very harshness of my voice served to warn me that something was amiss. Surely the child's responsibility in this matter was, at best, 10%. The remaining 90% belonged to me, but by denying my own role, someone else had to bear a greater burden. That is, denial of my own responsibility required that responsibility be strongly projected onto someone else, to balance the "responsibility equation."

In a somewhat similar fashion, it has been argued that denying one's own homosexual tendencies will cause one to project these sexual tendencies onto others. It is as if we are aware that there is some homosexual content in the immediate neighborhood and, denying our own portion, we go looking for the missing homosexuality in others. Some striking experimental work has recently been produced in support of this possibility.¹² Fully heterosexual men (no homosexual behavior, no homosexual fantasies) are divided into those that are relatively homophobic and those who are not. Homophobic men are defined as those who are uncomfortable with, fearful of, and hostile toward homosexual men. Homophobia is measured by a series of 25 questions (TABLE 2). A rough analogue with the GSR was provided by a plethysmograph attached to the base of the penis which measures changes in circumference, while interviews provided information on conscious perception of tumescence and arousal. Of course, we are unconscious of our GSRs, but conscious of changes in penile circumference, at least beyond some threshold, so the analogy is not precise, but the methodology provides results of parallel interest to those of Gur and Sackeim.⁶

When the two groups of men are exposed to four-minute sexual videos (heterosexual, lesbian, and male homosexual), the plethysmograph shows that both sets of men respond with similar levels of arousal to the heterosexual and lesbian videos but that only the *homophobic* men show a significant response to the male homosexual

video. Interviews afterwards show that both categories of men give accurate estimates of their degree of tumescence and arousal to all stimuli with one exception: the homophobic men deny their response to the male homosexual video!

The results make a certain kind of superficial sense. Those heterosexual males who are, according to their own account, fully heterosexual in behavior and in fantasy yet who will actually experience arousal to the sight of two men making love would be expected to be more uncomfortable in the vicinity of homosexual men. These men, after all, represent continual possible sources of arousal for the man's latent homosexual affect. Discomfort around homosexuals and disgust and anger at them may be expected to be larger where homosexual threat is greater. Note again a dynamic between denial and projection. Denying their own homosexual feelings may force the individual to project a greater danger of those same tendencies onto others.

Ramachandran^{10,11} has recently produced very striking evidence that processes of denial—and subsequent rationalization—appear to reside preferentially in the left brain. People with a stroke on the right side of the body (damage to the left brain) never or very rarely deny their condition, while a certain, small percentage of those with left-side paralysis deny their stroke and, when confronted with strong counter-evidence, indulge in a remarkable array of rationalizations denying the *cause* of their inability to move (arthritis, a general lethargy, etc.). This is consistent with other evidence that the right hemisphere is more emotionally honest, while the left hemisphere is actively engaged in self-promotion. It goes without saying that we need much more evidence on the underlying physiology, neurobiology and anatomy of mechanisms of self-deception.

INTERACTION WITH OTHER BIOLOGICAL SYSTEMS

Internal conflict and biased information flow within the individual probably have multiple biological sources, self-deception evolving in the service of deceit being only one. The alternative sources are taken up here with particular attention to their interactions.

1. Parent-offspring conflict. As is now well recognized, parents and offspring are expected to be in conflict in any outbred species since each will be related to self by 1 but to the other by only 1/2. This 1/2 degree of relatedness leads to a strong overlap in self-interest, but also an imperfect one, giving scope to various kinds of conflict.¹⁸ Especially important in our own species is the fact that parent-offspring conflict extends to the behavioral tendencies of the offspring with the parent being selected to mold a more altruistic and less selfish offspring—at least as these behaviors affect other relatives—than the offspring is expected to act on its own. On the assumption that an internal representation of the parental voice is valuable to the child when their interests overlap closely, it can easily be imagined that selection has accentuated the parental voice in the offspring to benefit the parent and that some conflict is expected within an individual between its own self-interest and the internal representation of its parents' view of its self-interest.

It is easy to imagine that mechanisms of deceit and self-deception could be parasitized in this interaction. For example, low parental investment may coexist with exaggerated displays of parental affections, the latter serving as cover for the former.

The offspring may be tempted to go along with the parental show since resistance and malaffection may lead to even less investment. Yes, mommy loves you and you love mommy too. But one can easily imagine that having to adopt this self-deception as one's own may have long-term negative consequences and may lead to later internal psychological conflict when you are no longer under your parent's immediate control (via investment).

A good clinical example of this is provided by Dori LeCroy.³³ A thirty-year-old woman arrived for therapy appointments in a hesitant and apprehensive manner which (when challenged) she explained by her desire to avoid intruding on another's "personal space." In a wispy, vacant style she described herself as a loving and "spiritual" person who put special value on kindness, tolerance, and forgiveness. She related events in her life with an emphasis on ill-treatment by friends and relatives, including physical abuse as a child from her alcoholic mother, but the complaints were accompanied by rationalizations which absolved others of blame (her mother was really "a beautiful person" with troubles of her own, for example). Most notably, she displayed no anger, no outrage, no desire for revenge. Instead, she worried about the well-being of the perpetrators! LeCroy speculates that abuse suffered as a child led to overidentification with the abuser: "Self-deception of this kind would have enabled her to behave devotedly as abused children frequently do, and thereby solicit nurture."³³ It is important to note that the woman had not forgotten the facts regarding the past, indeed she volunteered them, but she had apparently transmuted her anger and resentment into oversolicitous indulgence, first for her mother and then for others. To reconcile the facts of the matter—that the maternal abuse was just fine—she had to agree to a negative self-image: she became bad and her mother good. Recently her mother has come around and now provides some real investment, but the patient herself still seems saddled with an imposed self-deception going to the heart of her identity. We are not attracted to people with a negative self-image, too timid to intrude and displaying an otherworldly attachment to altruism, even in the face of mistreatment!

2. *Internal genetic conflict.* A stronger potential source of internal conflict and biased information flow within the individual is internal *genetic* conflict, due to differing degrees of relatedness to others enjoyed by different parts of an individual's genome; these in turn are due to different rules of inheritance. For example, mitochondrial DNA (mtDNA) is passed only mother to offspring, while, of course, the autosomes are passed from each parent. One kind of conflict this can set up is over inbreeding. Autosomes enjoy an increased relatedness to offspring when they practice inbreeding, but this is not true for mtDNA, which is always related to progeny on the maternal side by 1. Put another way, since the mtDNA in any given individual is only coming from one parent, it does not increase relatedness for that parent to be related to the second parent. An autosome deciding whether inbreeding would be advantageous has to set against the increase in relatedness a decrease in quality of the offspring due to inbreeding depression. But the mtDNA will only see the inbreeding depression; thus as long as there is any inbreeding depression (and there often is), mtDNA will oppose inbreeding that the autosomes may favor.

There is no evidence regarding such interactions in animals, but there is striking evidence of exactly this kind of conflict in plants (for a good review of the relevant theory, see Ref. 34). Since most plants are hermaphrodites they can, in principle, practice "selfing" where the pollen and the ovules come from the same plant: this

raises degree of relatedness to offspring from 1/2 to 1. About a 1/2 to 3/4 of all flowering plants are capable of selfing and in these species (but only rarely in obligately non-selfing species) one finds a most interesting conflict: mtDNA causes abortion of the male function or sterile pollen (cytoplasmic male sterility), while the nuclear DNA often acts to re-establish male function. This is exactly consistent with mtDNA's always opposing inbreeding when there is an outbred alternative.

The kind of conflict I have been describing pits a small part of the genotype against almost all the other genes: for this reason, such conflict is expected to be infrequent and resolved usually in favor of the dominant set of genes. A more important kind of internal genetic conflict for our own species pits one-half of the genotype against the other half. I refer to the phenomenon of genomic imprinting or parent-specific-gene expression (reviewed in Ref. 35). A small number of genes in us have the property that they are expressed only when inherited from one sex, the copy inherited from the opposite sex being silenced (or sometimes there is only a quantitative difference in gene expression depending on parent-of-origin).

The importance of genomic imprinting is that it allows imprinted genes to act on the basis of exact degrees of relatedness to each parent. This inevitably leads to conflict between paternal and maternal genes.³⁶ Possible psychological conflicts arising from imprinting are easy to describe.³⁷ Consider, for example, a contemplated act of inbreeding with your mother's sister's offspring. You are related on the maternal side and will thus enjoy an increase in relatedness to any resulting offspring by inbreeding on the maternal side, but the paternal genes will enjoy no increase in relatedness though they will suffer any inbreeding depression associated with the inbreeding. We can imagine your maternally active genes urging you to consider the inbreeding while your paternally active genes might take a moralistic posture and emphasize the biological defects thereby generated. Whether mtDNA has also been selected to decrease inbreeding, as in plants, is as yet unknown.

There is very intriguing indirect evidence suggesting that parts of the body may differ in the degree to which they express maternally active versus paternally active genes (reviewed in Ref. 19). In mice chimeras which consist of a mixture of normal cells with cells that have either a double dose of maternal genes (and no paternal ones) or a double dose of paternal genes (and no maternal ones), it turns out that the two added kinds of cells survive and proliferate differentially according to tissue: thus, doubly maternal cells do well in the neocortex of the brain but do not survive and proliferate in the hypothalamus and vice versa for doubly paternal cells. By similar logic the tissue producing dentin appears to be more maternally active, while the tissue producing enamel is more paternally active. Thus, it is possible that there are conflicts at the level of tissues in which one can also imagine *selves*-deception, that is, deceitful signals sent out from one tissue, overemphasizing one parent's interests whose signals are devalued by another tissue, overemphasizing the opposite sexed parent's interests. Where maternal kin are much more frequent in the social group than paternal kin, maternally active tissue in the neocortex may say, in effect, "Family is important, I like family, I believe in investing in family" while the hypothalamus may reply, "I'm hungry!"³⁷

We can imagine interactions between genomic imprinting and other systems we have been discussing. For example, parental indoctrination will work better when it interacts with the appropriate imprinted genes: maternal manipulation with maternally active genes in the progeny and paternal manipulation with paternally active

genes.¹⁹ At the same time it is easy to imagine that mechanisms useful in self-deception to deceive others may prove useful in within-individual conflict. If selfish impulses are kept unconscious, the better to hide them from others, and they may also stay unconscious, the better not to be spotted by oppositely imprinted genes.

3. *Selection to suppress negative traits.* Everyone can expect to have some negative traits that are stuck in the phenotype either through misdevelopment or through genetic defect, and these are likely to have been such a regular part of our existence for so long that we may well wonder whether selection has not favored a mechanism which searches for such negative traits and attempts to suppress them. All genes in the individual would be in agreement with such a program, including the defective gene. Mutation will inevitably supply some negative traits,³⁸ but it is well to be aware of the fact that even in the absence of such a supply some selective factors by themselves generate some negative traits. For example, sex antagonistic genes are those that have opposite effects on reproductive success when found in the two sexes. As long as the net effect is positive, the gene will be favored, even though, when found in the opposite sex, it has negative effects on lifetime reproductive success.

William Rice's^{39–41} beautiful experiments on *Drosophila* demonstrate clearly that sex antagonistic genes are a regular part of the *Drosophila* genome and by extension are expected in all sexually reproducing species that are not perfectly, life-long monogamous. This means that each sex is a partial compromise between the two sexes and contains numerous traits disadvantageous to that sex (but advantageous in the opposite sex).

Naturally, if a mechanism for suppressing negative traits does exist, one may well expect internal conflict, forces acting to maintain the negative trait being opposed by efforts at suppression. There is no selection to increase the resistance, but as suppression is selected to become more effective, more negative genes will remain in the genotype because the suppression has reduced or eliminated the cost. It is easy to imagine an interaction between this mechanism and parent-offspring conflict, since parents may help you locate—and encourage you to suppress—such negative traits, but due to imperfect overlap in self-interest, they may encourage you to think a trait negative to yourself when it is in reality only negative to themselves. Similarly, it is conceivable that paternally active genes (for example) may attempt to suppress maternally active ones (or vice versa) by pretending that it is an organism-wide negative phenotypic trait that needs to be suppressed.

Prayer and meditation are two widespread examples of people wrestling with their phenotypes, some of which may have been favored by selection to suppress negative phenotypic traits, including the negative phenotypic trait of self-deception! Many famous passages from the world's great religions, as well as rituals of prayer and meditation, are directed against self-deception, as in this loose translation of Matthew 7:1–5 in the New Testament of the Bible: “Judge not that ye be not judged, for you are projecting your faults onto others; get rid of your own self-deception first, then you will have a chance of seeing others objectively.”

4. *Positive illusions?* Another important possibility is that self-deception has intrinsic benefit for the organism performing it, quite independent of any improved ability to fool others. In the past twenty years an important literature has grown up^{8,9} which appears to demonstrate that there are intrinsic benefits to having a higher perceived ability to affect an outcome, a higher self-perception, and a more optimistic view of the future than facts would seem to justify. It has been known for some time

that depressed individuals tend not to go in for the routine kinds of self-inflation that we have described above. This is sometimes interpreted to mean that we would all be depressed if we viewed reality accurately, while it seems more likely that the depressed state may be a time of personal re-evaluation, where self-inflation would serve no useful purpose. While considering alternative actions, people evaluate them more rationally than when they have settled on one option, at which time they practice a mild form of self-deception in which they rationalize their choice as the best possible, imagine themselves to have more control over future events than they do, and see more positive outcomes than seem justified. What seems clear is that they gain direct benefits of functioning from these actions.⁴² Life is intrinsically future-oriented and mental operations that keep a positive future orientation at the forefront result in better future outcomes (though perhaps not as good as those projected). The existence of the placebo effect is another example of this principle (though it requires the cooperation of another person ostensibly dispensing medicine). It would be very valuable to integrate our understanding of this kind of positive self-deception into the larger framework of self-deceptions we have been describing.

SELF-DECEPTION AND HUMAN DISASTERS

There can be little doubt that self-deception makes a disproportionate contribution to human disasters, especially in the form of misguided social policies, wars being perhaps the most costly example. This is part of the large downside to human self-deception. Since the general cost of self-deception is the misapprehension of reality, especially social reality, self-deception may easily generate large social costs (everyone on the airplane dies, the entire nation is devastated by a war some of its members started).

Disasters are, of course, studied in retrospect so the evidence is not yet scientific for the connection to self-deception, but it is certainly suggestive. In the following examples, we also see how analysis of individual self-deception can easily be extended to groups: pairs of individuals, an organization and an entire society.

Two-party self-deception. Trivers and Newton's² analysis of the crash of Air Florida's Flight 90 suggests that the pilot was practicing self-deception and the co-pilot acquiesced. The first clue comes from the cockpit conversation during take-off (TABLE 3). The co-pilot was flying the airplane, yet it was he who noticed contradictory information from the instrument panel and repeatedly spoke while it mattered (i.e., while they could still safely abort the flight). The pilot spoke only once, offering a false rationalization for the disturbing instrument readings. Only when it was too late—they were in the air—did the pilot start talking, while the co-pilot fell silent. An analysis of their conversation prior to take-off showed a consistent pattern of reality denial by the pilot (TABLE 4). His casual approach to reality, coupled with overconfidence, may have served him well in many minor situations, but proved fatal when real danger required close attention to reality, including the psychological state of his co-pilot.

When an organization practices deception toward the larger society, this may induce organizational self-deception. Richard Feynman³ analyzed the cause of the Challenger disaster and concluded that NASA's deceptive posture toward U.S. soci-

TABLE 3. Crash of Air Florida flight 90²

	Co-pilot	Pilot
During take-off	Speaks	Silent ^a
After lift-off	Silent	Speaks

^aExcept for one rationalization.

TABLE 4. Conversations during taxiing prior to take-off²

Co-pilot	Pilot
<ul style="list-style-type: none"> • Detailed description of snow on wings • Calls attention to danger they face (too long since de-icing) • Asks for advice on take-off 	<ul style="list-style-type: none"> • Diminutive description of snow on wings • Deflects attention to ideal world (de-icing machine on runway) • Tells him to do what he wants

TABLE 5. Feynman's analysis of NASA's shift to self-deception³

1960s	Aim:	<ul style="list-style-type: none"> • Go to moon • No conflict with larger society • No internal conflict re facts • Built from bottom up
	Result:	<ul style="list-style-type: none"> • Success
1970s	Aim:	<ul style="list-style-type: none"> • Emphy a \$5 billion bureaucracy • Need to convince larger society — repeated manned flight via shuttle • Bottom splits from top, which does not wish to know true facts re safety • Built from top down
	Result:	<ul style="list-style-type: none"> • Challenger disaster

ety had bred organizational self-deception. When NASA was given the assignment and the funds to travel to the moon in the 1960s, the society, for better or worse, gave full support to the project: Beat the Soviets to the moon (TABLE 5). As a result, NASA could design the moon vehicle in a rational way. The vehicle was designed from the bottom up, with multiple alternatives tried at each step, permitting maximum flexibility as the spacecraft was developed. Once the U.S. reached the moon, NASA was a five-billion-dollar bureaucracy with no work to do. Its subsequent history, Feynman argued, was dominated by a need to generate funds, and critical design features, such as manned flight versus unmanned flight, were chosen precisely because they were costly. In addition, manned flight had glamour appeal, which would generate enthusiasm for the funding. At the same time it was necessary to sell this project to Congress and the American people. The very concept of a reusable vehicle—the so-called Shuttle—was designed to appear inexpensive, while in fact it was very costly (more expensive, it turned out, than using brand new devices each time).

Means and concepts were chosen for their ability to generate cash-flow and the apparatus was then designed top-down. This had the unfortunate effect that when a problem surfaced, such as had with the O-rings, there was little parallel exploration or knowledge to solve the problem. Thus NASA chose to minimize the problem and the unit within NASA that was consigned to deal with safety became an agent of rationalization and denial, instead of one of rational study of safety factors.

Some of the most extraordinary mental gyrations in service of institutional self-deception occurred within the Safety Unit. Seven of twenty-four Challenger flights had shown O-ring damage. Feynman showed that if you merely plotted chance of damage as a function of temperature at time of take-off you got a significant negative relationship: lower temperature meant a higher chance of O-ring damage. To prevent themselves from seeing this, the Safety Unit performed the following mental operation. They said that seventeen flights showed no damage and were thus irrelevant and could be excluded from further analysis. Since some of the cases of damage occurred during high-temperature take-offs, temperature at take-off could be ruled out as a causative agent. One of the O-rings had been eaten 1/3 of the way through. Had it been eaten all the way through, the flight would have blown up, as did the Challenger. But NASA cited this case of 1/3 damage as a virtue. They claimed to have built in a "threefold safety factor"! This is a very unconventional use of language. By law you must build an elevator strong enough so that the cable can support a full load with no damage. Then you must make it eleven times stronger. This is called an eleven-fold safety factor. NASA has the elevator hanging by a thread and calls it a virtue. They even used circular argumentation with a remarkably short radius: since manned flight had to be much safer than unmanned flight, it perforce was. In short, in service of the larger institutional deceit and self-deception, the Safety Unit was thoroughly corrupted to serve propaganda ends, that is, to create the appearance of safety where there was none.

There is thus a close analogy between self-deception within an individual and self-deception within an organization, both serving to deceive others. In neither case is information completely destroyed (all 12 engineers at Thiokol, which built the O-ring, voted against flight that morning). It is merely relegated to portions of the person or the organization that are inaccessible to consciousness (we can think of the people running NASA as the conscious part of the organization). In both cases the entity's relationship to others determines its internal structure of information. In a non-deceitful relationship information can be stored logically and coherently. In a deceitful relationship information will be stored in a biased manner the better to deceive others—but with serious potential costs. Note, however, that it is the astronauts who suffered the ultimate cost, while the upper echelons of NASA—indeed, the entire organization minus the dead—may have enjoyed a net benefit (in employment, for example) from their casual and self-deceived approach to safety.

Self-deception is especially likely in warfare. Richard Wrangham has recently extended the analysis of self-deception to human warfare in a most revealing way.⁵ Evolutionary logic suggests that self-deception is apt to be especially costly in interactions with outsiders, members of another group. In interactions with group members, self-deception will be inhibited by two forces: a partial overlap in self-interest gives greater weight to the opinion of others and within-group feedback provides a partial corrective to personal self-deception. In interactions between groups, everyday processes of self-enhancement are uninhibited by negative feedback from oth-

ers, nor by concern for their welfare, while derogation of the outsiders' moral worth, physical strength, and bravery is likewise unchecked by feedback and shared self-interest. These result in faulty mechanisms of assessment, and aggression will be more likely where each partner is biased in an unrealistic direction in self- and other-assessment, making conflicts more likely to occur and contests therefore more costly, on average, without any average gain in benefits.⁵ Derogation of the moral status of your enemies only makes you underestimate their motivation (consider U.S. assessment of the Vietnamese). For an excellent analysis of this phenomenon, as applied to the Old Testament of the Bible, see Hartung.⁴³

Processes of group self-deception only make matters worse: Within each group individuals are misoriented in the same direction, easily reinforcing each other and absence of contrary views is taken as confirming evidence (even silence is misinterpreted as support).⁵ Tuchman⁴ has frightening stories to tell of an individual leader and his cohorts whipping themselves into a frenzy of self-deception prior to launching an ill-advised, indeed disastrous, attack on neighbors.

Military incompetence—losing while expecting to win—is accompanied by four common symptoms: overconfidence, underestimation of the neighbor, ignoring intelligence reports, and wastage of manpower.⁵ The latter two are noteworthy. The logic of self-deception preserves conscious illusion by becoming unconscious of contrary evidence, even when provided by one's own agents, whose very purpose it is to provide accurate information. Note in the Challenger disaster how the unit assigned to consider safety ended up being subverted to rationalize unsafety, even though its ostensible purpose was to view the matter objectively.³ Wastage of manpower is a direct cost of self-deception since forces are deployed along illusory lines of attack, instead of rationally calibrated toward the real situation.

Wrangham makes an important distinction between raids and battles.⁵ Lethal raids are attacks on a few neighbors, with numerical superiority being a key stimulus to attack. Raids have a long evolutionary history (chimpanzee males practice lethal raids)⁴⁴ and opportunities for self-deception are minimized by the ease of rational assessment (e.g., evidence of numerical superiority). Battles are set pieces between large opposing armies. They are a recent invention (within historical times, more or less), rational assessment is much more difficult, and a long evolutionary history of derogating others¹³ makes misassessments especially likely. In short, we should be especially vigilant in guarding against self-deception when contemplating warfare.

CONCLUSION

Self-deception appears to be a universal human trait which touches our lives at all levels—from our innermost thoughts to the chance that we will be annihilated together in warfare. It affects the relative development of intellectual disciplines (the more social the content, the less developed the discipline: contrast physics and sociology) as well as the relative degree of consciousness of individuals (generally, more self-deceived, less conscious). An evolutionary analysis suggests that the root cause is social, including selection to deceive others, selection on others to manipulate and deceive oneself, and selection on competing sections of one's own genotype. There are undoubtedly complex and important interactions between these (and other) kinds of self-deception. The relevant evidence stretches from personal anecdote to histor-

ical analysis, but we especially need more biological evidence on the genetics, endocrinology, physiology and neuroanatomy of self-deception and we need to integrate very disparate findings from experimental, social and clinical psychology into the evolutionary analysis. We also need a detailed theory for the evolution of deception (many elements exist already) and a theory of consciousness based on our understanding of self-deception. Evolutionary theory promises to provide a firm foundation for a science of self-deception, which should eventually be able to predict both the circumstances expected to induce greater self-deception and the particular forms of self-deception being induced.

ACKNOWLEDGMENTS

I am grateful to Drs. Helena Cronin, Dori LeCroy, and Peter Moller for encouraging me to publish this article, to Mr. John Martin (Rockford, Illinois), and to the Ann and Gordon Getty Foundation for generous financial support. I am also thankful to Drs. David Haig, Dori LeCroy, David Smith, and especially Richard Wrangham for a series of detailed and valuable comments on the manuscript.

REFERENCES

1. MCEWAN, I. 1997. *Enduring Love*. Jonathan Cape. London.
2. TRIVERS, R.L. & H.P. NEWTON. 1982. The crash of Flight 90: doomed by self-deception? *Science Digest* (November) : 66, 67 and 111.
3. FEYNMAN, R. 1988. *What Do You Care What Other People Think? Further Adventures of a Curious Character*. Norton. New York.
4. TUCHMAN, B. 1988. *The March of Folly: From Troy to Viet Nam*.
5. WRANGHAM, R. 1999. Is military incompetence adaptive? *Evol. Hum. Behav.* **20**: 3–12.
6. GUR, R. & H. A. SACKEIM. 1979. Self-deception: a concept in search of a phenomenon. *J. Pers. Soc. Psychol.* **37**: 147–169.
7. GREENWALD, A.G. 1980. The totalitarian ego: fabrication and revision of personal history. *Am. Psychol.* **35**: 603–618.
8. TAYLOR, S.E. & D.A. ARMOR. 1996. Positive illusions and coping with adversity. *J. Pers.* **64**: 873–898.
9. TAYLOR, S.E. 1998. Positive illusions. *In Encyclopedia of Mental Health*, Vol. 3. H.S. Friedman, Ed.: 199–208. Academic Press, San Diego, CA.
10. RAMACHANDRAN, V. & D. ROGERS-RAMACHANDRAN. 1996. Denial of disabilities in anosognosia. *Nature* **382**: 501.
11. RAMACHANDRAN, V. 1997. The evolutionary biology of self-deception, laughter, dreaming and depression: some clues from anosognosia. *Med. Hypotheses* **47**: 347–362.
12. ADAMS, H.E., L. W. WRIGHT, Jr & B.A. LOHR. 1996. Is homophobia associated with homosexual arousal? *J. Abnorm. Psychol.* **105**: 440–445.
13. KREBS, D.L. & DENTON, K. 1997. Social illusions and self-deception: the evolution of biases in person perception. *In Evolutionary Social Psychology*. J. A. Simpson & D.T. Kenrick, Eds: 21–48. Erlbaum Associates. Mahwah, NJ.
14. HAMILTON, W. D. 1994. The genetical evolution of social behaviour. I, II. *J. Theor. Biol.* **7**: 1–52.
15. TRIVERS, R. 1985. *Social Evolution*. Benjamin/Cummings. Menlo Park, CA.
16. KREBS, J.R. & R. DAWKINS. 1984. Animal signals: mindreading and manipulation. *In Behavioural Ecology*, 2nd ed. J.R. Krebs & N.B. Davies, Eds.: 380–402. Sinauer. Sunderland, MA.

17. ALEXANDER, R.D. 1979. Darwinism and Human Affairs. University of Washington Press: Seattle, WA.
18. TRIVERS, R.L. 1974. Parent-offspring conflict. *Am. Zool.* **14**: 249–264.
19. TRIVERS, R. & A. BURT. 1999. Kinship and genomic imprinting. *In* Genomic Imprinting. An Interdisciplinary Approach. R. Ohlsson, Ed.: 1–23. Springer. Heidelberg, Germany.
20. LEWIS, M. 1997. Altering Fate: Why the Past Does Not Predict the Future. Guilford Press. New York, NY.
21. TIGER, L. 1979. Optimism: The Biology of Hope. Simon and Schuster. New York.
22. ROHWER, S. 1977. Status signalling in Harris sparrows: some experiments in deception. *Behaviour* **61**: 107–129.
23. ROHWER, S & F.A. ROHWER. 1978. Status signalling in Harris sparrows: experimental deception achieved. *Anim. Behav.* **26**: 1012–1022.
24. MØLLER, A.P. & J.P. SWADDLE. 1987. Social control of deception among status signalling house sparrows *Passer domesticus*. *Behav. Ecol. Sociobio.* **20**: 307–311.
25. TRIVERS, R. 1991. Deceit and self-deception: the relationship between communication and consciousness. *In*: Man and Beast Revisited. M. Robinson&L.Tiger, Eds.: 175–191. Smithsonian. Washington, DC.
26. MELE, D. 1997. Real self-deception. *Behav. Brain Sci.* **20**: 91–136.
27. FOSTER E, M. JOBLING, P. TAYLOR, P. DONNELLY, P. DE KNIFF, R. MIEREMET, T. ZERJAL & C. TYLER-SMITH. 1998. Jefferson fathered slave's last child. *Nature* **396**: 27–28.
28. CASHDAN, E. 1995. Hormones, sex and status in women. *Horm. Behav.* **29**: 354–366.
29. LIBET, B. 1996. Neuronal time factors in conscious and unconscious mental functions. *In* Toward a Science of Consciousness: The First Tucson Discussion and Debates. S. R. Hameroff, A.W. Kaszniak & A. Scott, Eds.: 337–347. MIT Press. Cambridge, MA.
30. DOUGLAS, W. & K. GIBBINS. 1983. Inadequacy of voice recognition as a demonstration of self-deception. *J. Pers. Soc. Psychol.* **44**: 589–592.
31. SACKEIM, H.A. & R.C. GUR. 1985. Voice recognition and the ontological status of self-deception. *J. Pers. Soc. Psychol.* **48**: 213–215.
32. HUDSON, W.W. & W.A. RICKETTS. 1980. A strategy for the measurement of homophobia. *J. Homosexuality* **5**: 356–371.
33. LECROY, D. 1998. Darwin in the clinic: an evolutionary perspective on psychodynamics found in a single case study. *ASCAP* **11**: 6–12.
34. FRANK, S. 1989. The evolutionary dynamics of cytoplasmic male sterility. *Am. Naturalist* **133**: 345–376.
35. OHLSSON, R. Ed. 1999. Genomic Imprinting. An Interdisciplinary Approach. Springer. Heidelberg.
36. HAIG, D. 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proc. Roy. Soc. Lond. B.* **264**: 1657–1662.
37. TRIVERS, R. 1997. Genetic Basis of intra-psyche conflict. *In* Uniting Psychology and Biology: Integrative Perspectives on Human Development. N. Segal, G.E. Weisfeld & C.C. Weisfeld, Eds.: 385–395. Am. Psychol. Assoc. Washington, DC.
38. EYRE-WALKER, A. & P.D. KEIGHTLEY. 1999. High genomic deleterious mutation rates in hominids. *Nature* **397**: 344–347.
39. RICE, W. 1992. Sexually antagonistic genes: experimental evidence. *Science* **256**: 1436–1439.
40. RICE, W. 1998. Male fitness increases when females are eliminated from gene pool: implications for the Y chromosome. *Proc. Natl. Acad. Sci. USA* **95**: 6217–6221.
41. HOLLAND, B. & W.R. RICE. 1999. Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proc. Natl. Acad. Sci. USA* **96**: 5083–5088.
42. KREBS, D., K. DENTON & N.C. HIGGINS. 1988. On the evolution of self-knowledge and self-deception. *In* Sociobiological Perspectives on Human Development. K. B. MacDonald, Ed. Springer. New York.
43. HARTUNG, J. 1995. Love thy neighbor: the evolution of in-group morality. *Skeptic* **3**: 86–99.
44. WRANGHAM, R. & D. PETERSON. 1996. Demonic Males: Apes and the Origins of Human Violence. Houghton Mifflin. Boston, MA.