Endnotes

Trivers would name this behavior Self-deception as self-promotion, as we shall see.

An earlier version of this paper was read at the meeting of the International Society for the History, Philosophy, and Social Science, of Biology in Vienna, Austria, July 2003. I would like to extend my gratitude to the members of my section for their valuable comments. I would also like to thank The Konrad Lorenz Institute, ISHPSSB, the NSF, and the Werkmeister Foundation for financial support. Also thanks to Zac Ernst, Al Mele, Robert Trivers, and Jason Zinser for help ful comments and criticisms on the final version.

References

Mele, Alfred. 2001 *Self-Deception Unmasked* .Princeton NJ: Princeton UP.

Trivers, R.1985 *Social Evolution* .Menlo Park,Calif. Benjamin/ Cummings .

Trivers, R.2002 *Natural Selection and Social Theory*, NewYork, Oxford UP.

Trivers, R. &H.P.Newton.1982 "*The crash of Flight90*, doomed by sel f-deception?" Science Digest (November):66,67and111.

McErlean,J.2000 *Philosophies of Science* :Wadsworth.

Ruse, M.2000 *The Evolution Wars* , New Brunswick, NJ: Rutgers UP.

convince her son that broad shoulders and a positive mental attitude are unattractive to women, presumably for her own selfish reasons. He may believe what she says, and consciously attempt to de-accentuate those features by wearing baggy clothing ,and acting pessimistically. Unconsciously, though his genes keep his mother from ruining his life; when not conscious of the effort, he occasionally tips his hand by flexing his muscles or flashin a smile at ladies. His genes force him to know the truth, while his mother has forced him to believe a falsehood. We can not fight our genes!

6.2. Reply. Does the previous thought-experiment count as a case of satisfying Sackeim and Gur's necessary and sufficient conditions? It seems to be: he believes p and believes ¬ p simultaneously; he is not a ware of the true belief; and he is motivated to believe the word of his mother. Does it also satisfy Mele's sufficient conditions for self-deception? Again, it seems to: p is false; he is motivationally biased to believe p; and one could argue that the genetic data that he possesses provides better reason to believe ¬ p than p. Might we argue, as is possible in the case of Larry, that he holds one belief, then comes to hold another, while never holding the two simultaneously? I don't think so. Larry's coming to falsely believe ¬ p is part of a causal chain, that began with his belief in p. In the case of Bo, there is no causal interaction between the two beliefs. While one belief is caused by social factors, the opposite belief is determined by the genotype.

7.Conclusions, and Calls for Further Work

I've presented what I take to be a demonstration of the dual-belief condition, which simultaneously satisfies Mele's sufficient condition for entering self-deception, and is consistent with an evolutionary explanation for self-deception. Further work might take the results to the next level, including identifying actual empirical evidence to support my thesis, to show that this class of cases actually exists, rather than remaining a mere plausibility. I believe actual empirical evidence of the dual- belief condition might be founde. g., in cases in which strict adherence to religious principles conflict with genetic signals received during the onset of puberty. I leave further empirical work to my colleagues in the social sciences.

of some property to make it seem less obvious.

(4) The construction of biased social theory. We all have theories regarding our relationship to other people and to society. In these situations, one deceives one self about the nature of those relationships.

(5) Fictitious narratives of intention. In this final situation,one deceives one's self about then a ture of one's intention. I would like to go to C, but can't justify the travel. I'm happy to take the opportunity to go to B. However, once I'mat B, I can easily justify the small extra travel to C, but do not think of C until I'm at B. Our patterns of motivation may run deeper, remaining unconscious for along period of time.

What exactly do I claim to prove with these examples? Surely, none of the above situations require the dual-belief condition. Rather, the theme we find running throughout Self-deception in the Service of Deceit is the creation of self-serving motives which run just below the surface of the conscious mind. The dual-belief condition would be impossible, though, if it were not for the existence of some fragmentation of the mind one section of which is actively engaged in deception, while the other section is engaged in honorable activities. This leaves one only to show the relationship between two such activities, in order to demonstrate an empirical example of the dual-belief condition.

6.1. Empirical Studies. Self-deception in the service of Deceit is probably only one source of internal (conscious v. unconscious) conflict, and biased information flow. Other sources show conflict on the level of the genes, as I will demonstrate now.

By way of analogy, Trivers discusses internal conflict within the genes, which do seem to demonstrate the dual-belief condition (on the genetic level).Experiments on Drosophila show that sex antagonistic genes are part of the genome of all Drosophila, and therefore are expected by mates. The genome, then, simul-taneously attempts to disguise the antagonistic gene, and display it. Thus, "if mechanism for suppressing negative traits do exist, one may well expect ... forces acting to maintain then negative trait being opposed by efforts at suppression."(Trivers 2002:284-85) By extension, we see clear examples of internal conflict in the mind.Consider Bo. Bo's mother may

An obvious follow up question would be this: is it the case that I don't deserve these jobs (p) inconsistent with someone should give me a chance (¬ p)?Is this really a clear contradiction? I can only say that it seems that one can derive no one should give me a chance from I don't deserve these jobs, or I don't deserve these jobs from no one should give me a chance. I grant that some logical work would need to be done there to show an entailment relation between these two beliefs.

6. A New Demonstration

In his 2002, Trivers describes self-deception as an internal fragmentation and conflict.(Trivers 2002:273-77) While the true and false information is both stored in the mind, the organism has a bias toward storing the true information unconsciously, and the false consciously. This way of organizing knowledge, he claims, has the outside observer in its interest: the observer first spots the false information, and only later may come to know the true information, which is concealed in the mind of the deceiver. Trivers calls this' Self-Deception in the Service of Deceit. "We can expect to find it in the following five situations.

(1) Denial of ongoing deception. This is the standard case as described in his Social Evolution, and described above, in which one comes to believe what is false, so as to convince some other individual of the falsehood. A deception counts as ongoing, if the other individual is currently standing in front of you, or you are otherwise currently actively carrying on the deception.

(2) Unconscious modules involving deception. In this situation, the dominant activity is honest, but a minor activity is deceitful. The author describes a simple habit he has of stealing chalk during lectures, quite unconsciously, from himself, leaving himself without any. He further admits to unconsciously stealing all sorts of little things while his conscious mind is occupied with other matters. What is the benefit of keeping this unconscious? Not only can he act surprised if caught, but also it frees the mind to carry on with the other activities.

(3) Self-Deception as Self Promotion. These cases involve exaggerating some property to make it seem more prominent, or denial

in order to attract amore successful mate. Although Mele is correct when he says that if holding true beliefs unconsciously makes agents less successful deceivers than those holding no true belief (because of subtle physiological indicators), and if evolution selects for successful deception, then a more successful deceiver would be one who does not hold a simultaneous true belief.

Yet here is Larry. Larry does not disprove Mele's thesis. It is true that holding a true belief makes on eless successful at deception. But it is also true that having a false belief while simultaneously unconsciously holding the true belief makes one a better deceiver than one with no false belief at all. Until it is shown that Trivers's evolutionary explanation of self-deception is conceptually impossible, It should not be discarded out of hand.

5. Reply

How might Mele reply? He might reply saying that I'm confusing the fact of evolution with the theory of evolution.(Ruse, 2000:4-5) He might say that I'm attributing what only amounts to a theory as an empirical fact. My reply is only that my thesis is much weaker than that: Probably Trivers' explanation does not result in an empirical fact, and so does not meet Mele's challenge from SDU, Chapter 4. But as a viable theory, it ought not be discarded prima facie.

Why should an opponent be persuaded that Larry believes p while also believing ¬ p? It would be nice to show a relationship between the two attitudes, and therefore the two beliefs, such that maintaining one is causally sufficient for holding the other. Mele might reply that, like in so many other cases first Larry had a true belief, then he had a false belief, but never held the two simultaneously. This is a stronger objection, and the reply comes much less easily. I offer the following heuristic argument to show that as long as Larry holds the first belief, he will continue to hold the second (all things being equal).I point to the story and ask Larry, do you think you'll be successful? Do you think you deserve these jobs you are applying for? No, I don't deserve them. Then why are you getting frustrated? To which he would reply, because eventually someone should give me a chance. At least it sounds like Larry holds the two beliefs simultaneously.

evolved ... because natural selection favors ever subtler ways of deceiving others, "does not depend for its plausibility on the thesis that the agent simultaneously holds Bel (p) & Bel (¬ p .But I believe that in many cases, it does depend for its plausibility on the dual-belief condition. What Mele shows is that in many cases, including the example I give above, the agent might be no worse off, and sometimes better off, if he does not also hold the contradictory true belief p. In at least one class of cases, how ever, this does not seem to be the case. Consider the following example.

Larry knows all along that he will never be successful but wants to attract a wife. He believes that the best family situation would be for him to marry a woman who would be a success. But he also knows that, for some reason, and much to his disappointment, successful women are only attracted to successful men. Larry decides that the best strategy is just to act as if he will someday be successful, hoping that projecting the attitude will be sufficient to fool his prospective mate. He begins applying for corporate jobs for which he is not qualified, applying to professional programs to which he is not competitive, and hanging out at expensive singles bars. At first he neatly tucks all of his rejection letters away in a file, and continues the applications. In the mean time, he meets several successful women.

After several months, his rejection letter pile starts to grow quite large, and he begins to grow some what frustrated. I know I don't deserve any of these jobs, he thinks, but eventually someone should just give me a chance. He begins to apply more vigorously, believing that if he gets enough applications out, some one must eventually take a chance on him. This new attitude comes out in his conversations with women, and since he comes to believe it him self, he does not exhibit any of the tell-tale physiological signs on one trying to deceive a lady in a bar. Al though Larry believes that he will never be successful and that he does not deserve a chance (p), due to this belief he comes to believe that someone should just give him a chance ¬ p.

Evolution has selected for women who can better tell when men are lying to them in bars, and therefore subsequently making them better at choosing a mate. Evolution has made men like Larry able to convince themselves of a false belief, even when they know the truth,

ity to detect my deception, evolution provides me with increasing deceiving ability. It works in the following way.

Others identify physiological indicators of my deception. I need to try to hide or avoid those indicators. What would work better than if I believed that very thing which was false (¬ p)? With the ability to unconsciously believe ¬ p, I am able to avoid the resulting physiological indicators correlated with it. I am therefore in a better situation to deceive others if I am able to simultaneously deceive myself.(SDU:89, Mele cites Trivers:415-16.)

"Of course it must be advantageous for the truth to be registered some where," Trivers explains, "so that mechanisms of self-deception are expected to reside side by-side with mechanisms for the correct apprehension of reality."(Trivers 1985:416) Trivers sees the mind as structured such that it is split into public and private portions, and such that the interactions between the two portions are quite complex.

Mele replies, stating that Trivers' theory in no way requires the dual-belief condition in order to be considered plausible. It is just as plausible, Mele asserts, that,"self-deception that satisfies the set of sufficient conditions... offered in Chapter 3 without satisfying the dual-belief requirement is no less effective a tool for deceiving others.(Mele 2001:89) "Though Trivers' account relies on my false belief (¬ p), it does not, at the same time, rely on my simultaneously holding the true belief (p) .Take the following evolutionarily relevant case, for example. I believe that it will help my chances of getting a mate and perpetuating my genes if I convince women that I will some day be a successful doctor. Although my test scores and undergraduate grades are sub-standard, I nonetheless convince my self that I have a good chance at gaining entrance to, and successfully completing, medical school. In deceiving my self, I avoid those tell-tale physiological indicators of deception. In deceiving my self, I am that much more successful at deceiving others.

4. TheAnalysis

Trivers claims that there are possible in stances in which we might start with a belief p, then come to believe ¬ p. He gives this as a sufficient condition for self deception. All Mele has to show is that Trivers' evolutionary explanation, such that self-deception, "has

where as Mele only gives sufficient conditions. Second, in compliance with Mele's challenge, they argue for the view that an agent can simultaneously hold two contradictory beliefs.

It may be important to first discuss Sackeim and Gur's experiment involving voice recognition, not only to demonstrate one direction such empirical evidence might take, but also because Trivers cites further similar evidence involving homophobic men. I'll mention this briefly before moving on to a discussion of the evolutionary explanation provided by Trivers.

In Sackeim and Gur's experiment, subjects are played audio recordings of their own voices. Although they state that the voice they hear is not their own, certain physiological responses seem to indicate other wise. Galvanic skin response, for example, seems to indicate that the agent does indeed recognize his/her own voice from the audio sample. So at the same time the agent indicated that he believes that the voice is not his own($\neg$ p), his physiological response indicates that he knows that it is, in fact, his voice (p).

If, in fact, this is how the story goes, then we seem to have an instance of an agent simultaneously holding contradictory beliefs (bel p &bel $\neg$ p).There is some question as to whether an unconscious galvanic skin response is sufficient for a belief. And in fact, Sackeim later admits this fact.(Mele 2001: 132, Sacheim1998:161-62)

While this particular experiment seems to have failed, it does not close the door on the possibility of the empirical instance of self-deception, nor on the Sackeim and Gur model. Before closing Chapter 4, Mele discusses the research of the biologist and anthropologist Robert Trivers.(Meler 2001:89)

3. Trivers' Social Evolution

Robert Trivers claims to find convincing evidence for self-deception that accords With Sackeim and Gur's conception. According to Trivers, evolution endorses an increasingly subtler self-deceiving ability. As our evolutionary need to deceive others increases, so must our ability to match that need. Others might recognize that subtle changes in my physiology (e.g., galvanic skin response, stammering speech, or neglecting to make eye contact) might indicate that I'm attempting to deceive them. To circumvent the others' growing abil-

including this clause as part of a sufficient condition, and allows the details to come out in the analysis

(4) The data possessed by S provides greater reason to believe ¬ p than p. This

Is just to say that S's cognitive peers (i. e., other agents with a similar level of intelligence ,but lacking the requisite bias) would believe ¬ p, given the same data.

These four conditions, when taken jointly, then, provide a sufficient condition for self-deception. I tmight be argued that some (or possibly all) might be taken as necessary conditions for entering self-deception. This is an interesting question best reserved for another discussion, and in no way is Mele committed to this view, nor is such a view required to get his account off the ground.

2. The Dual-Belief Condition

In Chapter 4 of SDU, Mele discusses several empirical demonstrations of self- deception. On this conception, some psychologists have attempted to show that an agent can be self-deceived and require that the self-deceived agent simultaneously believe p and believe p. On the face of it, at least, this is a conceptual possibility: holding these two beliefs does not in itself entail a contradiction. Strictly speaking, there is a clear distinction between,

(1)[Bel (p & ¬ p)],and

(2)[Bel (p) & Bel (¬ p)].

Where only (2) describes the dual-belief condition. It would be much harder to satisfy condition (1), where the agent simultaneously believes (p & ¬ p).

Sackeim and Gur are one such pair of psychologists who take Mele's challenge to provide empirical evidence of self-deception, so described. They propose the following set of necessary and sufficient conditions for self-deception.(cf. Mele 2001:81)

(1)The agent simultaneously believes p and believes ¬ p.

(2)The agent is not a ware that he/she holds one of the beliefs.

(3)The determination of which belief is held is due to a motivationally biased act.

These conditions, obviously, differ from Mele's in several ways. First, they are intended to betaken as both necessary and sufficient,

An Evolutionary Explanation of Self-Deception

Robert C. Robinson

Abstract

In Chapter 4 of his *Self-Deception Unmasked* (hereafter, SDU), Al Mele considers several(attempted)empirical demonstrations of self-deception.

These empirical demonstrations work under the conception of what Mele refers to as the "dual-belief requirement, "in which an agent simultaneously holds a belief panda belief ¬ p. Toward the end of this chapter, Mele considers the argument of one biologist and anthropologist, Robert Trivers, who describes what he takes to be an evolutionary explanation for coming to form false beliefs. Mele argues briefly that Trivers's account is no more explanatory than a similar one that does not include the dual-belief requirement. I present a case describing Trivers' analysis, show how Mele might reply to it. After briefly explaining Mele's sufficient conditions for entering self-deception from Chapter 3 of SDU, I'll consider what it means to hold the dual-belief .I'll then consider what I take to be a class of cases of self-deception which rely on genetic determinism, which I take to satisfy the dual-belief condition

1. Mele's Sufficient Conditions

Mele claims that the following four conditions are jointly sufficient for S entering self-deception in acquiring a belief p.(Mele 2001:50-51)

(1) p, which S believes, is false. This is merely a semantic point, in that one cannot be deceived in believing p, and thus cannot be self-deceived in believing p, unless p is false.

(2) S is motivationally biased in his/her treatment of data relevant (or seemingly relevant) to the truth value of p .Mele describes several ways of entering self-deception in Chapters 1 & 2,where each require some motivational biasing of data.(ibid.:11-24)

(3) The bias is a non-deviant cause of S believing p .A discussion of causation in any sphere should include treatment of the problem of deviant causation. Mele hopes to avoid this discussion here by