

SELF-DECEPTION AS A WEIGHTLESS MASK

UDC 159.923.2

Elias L. Khalil

Monash University, Department of Economics, Clayton, Australia

Abstract. *Self-deception cuts across the behavior of persons, organizations, and states. Self-deception amounts to wearing a weightless mask: While the wearer is unaware of it, spectators usually recognize it. As soon as the wearer becomes conscious of the weightless mask, self-deception ceases to exist. This gives rise to a paradox: How could it ever be possible for the person to succeed in hiding the weightless mask from his or her own conscious self? To solve the paradox, this paper proposes that self-deception involves “two structural choices”: 1) an agent identifies the optimal decision — but the agent may fail (as a result of temptation) to choose it, which gives rise to self-blame; 2) The agent under focus may choose to invent a misleader (a lie) to make the choice appear “as if” it were optimal in order to avoid self-blame. The quest for a blameless choice makes it possible for the person to succeed in hiding the weightless mask from the conscious self. Aside from solving the Self-Deception Paradox, the proposed two-structural choice theory sheds light on three kinds of self-deception: red herring, self-rationalization, and denial.*

Key words: *rational choice, self-blame, temptations, image-management, just world hypothesis.*

1. THE SELF-DECEPTION PARADOX

In Aesop’s fable, “the fox and sour grapes,” the fox tries a couple of times to reach the grapes. Once the fox fails, the fox walks away telling itself a sudden excuse: “They must be sour grapes.” People, organizations, and states often use similar sudden or *ad hoc* excuses to rationalize laziness, myopic actions, inattentiveness, or transgressions of ethical boundaries. While such excuses allow actors to deceive themselves, they generally fail to deceive spectators. In Aesop’s fable, to start with, there are no spectators. This should suggest, at least *prima facie*, that the spectator is not a central feature of self-deception. What is central, as Aesop’s fable illustrates, in the definition of self-deception is the “self”:

Received November 13, 2015 / Accepted July 10, 2016

Corresponding author: Elias Khalil

Monash University, Department of Economics, Clayton, Vic 3800, Australia

E-mail: khalil@monash.edu

Definition: Self-deception consists of a person who deceives his or her own self. This amounts to a person with an intention to deceive, while the objective of deception is the “conscious self” of the person.

It is crucial that the “conscious self” is kept blind of the person’s intention. If the person’s intention is transparent to the conscious self, self-deception could not exist. But given that the deceiving and the deceived entities are constituents of the same person, self-deception can be described as “willful blindness.” The deceiving entity hides from the conscious self the full information that informs a choice. To keep the conscious self in a state of blindness of such information, the deceiving entity invents a misleader, i.e., a lie, an *ad hoc* belief, or an *ad hoc* interpretation of the facts. Misleaders assist the self-deceiving agent to avoid looking in the mirror and see the weightless mask behind which he or she hiding the discomforting facts — whereas the weightless mask is usually visible to the spectators.¹

But willful blindness gives rise to a paradox:

The Self-Deception Paradox: How could it ever be possible for the person to succeed in hiding his or her own intention from his or her own conscious self?

The Self-Deception Paradox has a neurobiological aspect, which falls outside the scope of this paper. On the other hand, the Self-Deception Paradox has a behavioral-philosophical aspect, which is the focus of this paper. Namely, how could a deceiving person ever succeed in hiding an intention from his or her own conscious self? This is a paradox given that the information could be available to the conscious self at no cost, while the deceiving self must expend nontrivial effort to obstruct the flow of such costless information (Taslit 2005; Hefferman 2011). So, the question from the behavioral-philosophical aspect, is what is the motive or incentive that is substantial enough to make it worthwhile for the deceiving self to expend the nontrivial effort to distort freely available information?

One immediate answer is the multiple-self theory (Elster 1986). The person is supposedly made up of two warring selves. Similar to any war, one party (self) tries to deceive the other. But there is one major problem with the multiple-self theory. In the case of war or conflict among separate individuals, each individual expressly tries to maximize his or her objective function and, hence, we have as many objective functions as there are individuals involved in the conflict. In the case of self-deception, there is only one objective function, viz., the one that expresses the wellbeing (utility) of a single individual. From common sense, which also expresses the basic pillar of rational choice theory, the individual acts in order to maximize the objective function, i.e., his or her utility. So, it is paradoxical as to why the individual would try to mislead him- or herself, i.e., try to undermine one’s utility. To state the Paradox in other words:

The Self-Deception Paradox--Rephrased: How could self-deception be possible in the sense that the individual tries to undermine his or her own wellbeing (utility)?

¹ The vagueness of some facts certainly makes it easier to invent misleaders and, hence, for the agent to deceive the self. However, vagueness is *not* a necessary condition of self-deception. Vagueness can only add a secondary consideration. It is easy to show, in a model, how self-deception is a positive function of vagueness: the greater is the vagueness, the lower is the cost of mental resources needed to implant and guard the misleader.

Evolutionary psychologists (von Hippel and Trivers 2011) try to solve the Self-Deception Paradox. They advance an almost convincing theory that has become most dominant one in the current literature on self-deception. They argue that agents who are involved in self-deception are actually maximizing their utility — i.e., they are not undermining their utility as supposed by the Self-Deception Paradox. The core condition of agents, for von Hippel and Trivers, is to try to deceive each other. The main focus of agents, for them, is for each one to deceive the other in a state of anarchy, conflict, and war. For the deception to succeed, it must be hidden from the other agent. But when agents try to hide their deception of others, it is too often that biological symptoms (such as sweating and nervousness of voice) betray their deceptive tactics, which undermines the deception. So, according to von Hippel and Trivers, agents must try to deceive themselves, and if successful at self-deception, they can suppress the biological symptoms. Once agents succeed in suppressing the biological symptoms, there is a much greater chance in deceiving others. That is, according to them, one misleads the self as a tactic to maximize one's utility.

The theory of von Hippel and Trivers has many problems. Given space limitations, this paper discusses only one problem (see Khalil 2015a). Why should, in the first place, anyone suffer from biological symptoms, such as sweating, when one is deceiving others? If there is a state of anarchy and war, it is legitimate to be a deceiver. After all, it is legitimate to try to rob, exploit, or even eat each other in the state of anarchy and war. If biological symptoms arise from lying, the state cannot be a state of anarchy and war. The state must rather be one of cooperation, reciprocity, and mutual respect of rights. If so, the lying or deception amounts to an aberration from an implicit or explicit social contract that assures cooperation. And such an aberration, given that the social contract expresses the optimal institution, amounts to suboptimality, i.e., acting contrary to one's non-myopic interest. So, the deception of others, the rock upon which von Hippel and Trivers' theory rests, cannot be optimal when biological symptoms are involved. If such deception is suboptimal, self-deception could never be optimal. Hence, we are back to the Self-Deception Paradox.

There are many other attempts, even less successful, to solve the Self-Deception Paradox. This paper cannot discuss them (Khalil 2015a). To solve the Paradox, we must first identify the benefit that the deceiving self seems to seek when one is involved in self-deception. Although this benefit may not justify the full cost of self-deception, it must be strictly positive given that self-deception involves nontrivial effort needed to buttress and defend misleaders. Otherwise, if there were zero benefit, there can be no reason or justification for self-deception *at all*. To understand the benefit of self-deception, we must start with decision theory, and particularly with rational choice theory.

The rationality starting point, though, can be puzzling given the Paradox, i.e., self-deception is non-rational. But the puzzle vanishes when we depart somewhat from rational choice theory. Such a theory entails that one makes a single choice or identifies a set of optimal choices, which would make self-deception appear as rational. Instead, this paper proposes that, in self-deception, one makes "two structural choices." First, one identifies the optimal decision, but takes a suboptimal action. Second, the deceiving self tries to cover the suboptimal act behind a weightless mask so that it appears to the conscious self "as if" it is optimal (rational).

While the proposed two-structural choice theory solves the behavioral-philosophical aspect of the Self-Deception Paradox, it also provides tools to distinguish among three kinds of self-deception: whether the decision maker is trying to rationalize a suboptimal action, to rationalize a suboptimal belief, or to rationalize the suboptimal search time of relevant facts.

2. TWO STRUCTURAL CHOICES

To solve the Self-Deception Paradox, this paper proposes the following core thesis:

Core Thesis: Granted the Self-Deception Paradox, i.e., given that self-deception is non-rational, self-deception paradoxically indicates that people, at a deeper level, are rational. So, a rational choice approach is imperative to explain self-deception.

This paper proposes the following modification of the rational choice approach. Namely, it proposes that self-deception consists of *two structural choices*:

First Structural choice: The agent succumbs to a temptation, i.e., chooses a suboptimal (non-rational) action. The action can be contrary to a commitment not to over-eat, over-sleep, under-save, and so on. The action can also be contrary to a commitment that is clearly defined as ethical, such as not to purchase goods from a company involved in abhorrent practices, to observe the privacy of others, and so on. This paper does not make a distinction between these two actions because the commitment not to over-eat or over-sleep is also ethical. After all, it is a commitment to a future self.

Second Structural choice: As a result of the first structural choice, the agent experiences awkwardness, self-blame, shame or emotional harm. These feelings are the outcome of the agent's own *ex ante* judgment that the agent is about to commit a suboptimal act — which means the agent can identify what is rational as opposed to the non-rational. The agent can either let him- or herself experience the deserved punishment of self-blame, which may result in an injury to self-concept and self-image, or can instead nullify such punishment by resorting to self-deception.

If self-deception is chosen, it entails the two choices. First, one has committed a suboptimal act. Second, one was successful in hiding behind the weightless mask, i.e., inserting the misleader and reconstructing the first choice away from the conscious self. The misleader consists, as broadly defined here, of inserting an input into the decision making which can be either an irrelevant fact or the unwarranted intervention of a new valuation of previous facts. With the insertion of a misleader, the suboptimal act looks to the conscious self as if it were an optimal one.

This proposed approach affirms the position of Fingarette (2000), which can be traced back to Sartre (1948). Namely, self-deception involves dishonesty or bad faith. But, on the other hand, there is a silver lining in such dishonesty. For people to suppress self-knowledge concerning their non-rational (suboptimal) act, it indicates that they care deeply about being rational. So, ironically, we need rational choice theory to dissect self-deception, that is, doubly non-rational choices.

While the two structural choices violate the axioms of rational choice, they violate them differently. First, succumbing to temptation means that action does not necessarily reveal one's optimal decision. This is a violation of the weak axiom revealed preference. Second, the manipulation of the facts to justify an actual choice means that the choice can cause the facts. This is a violation of the core definition of rational choice: causality must run from facts to decision, rather than the other way around.

But why do agents undertake the second structural choice, i.e., attempt to appear rational, when they are not? What is the benefit of self-deception? The apparent benefit is the negation of the ill-feeling of blame that accompanies the first structural choice in order to sustain "self-concept," "self-esteem," "symbolic utility," or "integrity." That is, the benefit consists

of what psychologists call “impression management” or “image management” (Aktipis 2000; Kurzban and Aktipis 2007). In effect, to sustain symbolic utility, the agent may choose to suppress, i.e., resort to self-deception, the negative emotion of self-blame that arises from succumbing to temptation.

The desire to sustain symbolic utility, and hence reduce self-blame, should be distinguished from the standard view of self-deception in economics and social psychology. Akerlof and Dickens (1982), Rabin (1994), and Konow (2000) postulate that self-deception is an act to reduce the “disutility” that arises from “cognitive dissonance.” As defined by the originator of the concept, Festinger (1957), cognitive dissonance arises when there is a discrepancy between one’s actual action and one’s principle or ethical commitment to how one should act. The discrepancy can involve an inconsistency between two upheld beliefs. The discrepancy gives rise to psychological disutility or tension, where “self-deception” is supposedly a way to bridge the discrepancy.

The existence of conflicting beliefs, i.e., cognitive dissonance, is an aspect of self-deception as defined here. However, cognitive dissonance, in itself, cannot capture self-deception. Why should the “bridging” of two inconsistent beliefs involve self-deceptive? Why is such bridging non-transparent to the self — as the self usually knows when it reconciles between altruism towards others and the pursuit of self-interest? There must be something that cannot be reconciled, in the first place, when one resorts to self-deception. That is, the actual choice and the contrary commitment must be irreconcilable to the point that one uses self-deception, i.e., a justification that is non-transparent to the self. Put differently, to characterize self-deception as a “bridge” designed to reduce cognitive dissonance begs the question: why should the bridging between action and commitment be hidden from the self?

This means that the theorist needs to go beyond the literature on cognitive dissonance to capture self-deception. The theorist should identify, first, the source of the discrepancy between the actual act and what one should do (Khalil 2000). Self-deception, contrary to the cognitive dissonance literature, cannot be seen as abridging between two legitimate ends: material utility as expressed by the actual act and some ethical utility as expressed by the commitment. Rather, one of the ends, viz., material utility, must have already been deemed, by the first structural act, as suboptimal.

The proposed two-structural choice theory has many payoffs. The most important one is the distinction among three kinds of (irrational) self-deception: red herring, self-rationalization, and denial. This requires us to discuss, first, the three layers of rational decision.

3. THREE LAYERS OF RATIONAL DECISION

Let us examine a decision such as the purchase of a beef sandwich. The purchase, as simple as it is, depends on given facts that include beliefs. For example, what is one’s belief concerning the restaurant hygiene? What is one’s belief concerning the health dimensions of beef? Beliefs, in turn, depend on other given facts, viz., information available to the consumer. And information, in turn, depends on other given facts such as the cost of search and the expected value of additional information. In short, the agent, when the consumer decides to consume X, is making three kinds of decisions: i) what is the optimal information to collect, given its effective cost; ii) what is the optimal belief, given the collected information; and

iii) what is the best bundle of goods to purchase, given beliefs. Hence, it is important to distinguish among the three layers of rational decisions:

- 1. Rationality of Consumption:** When agents make decisions concerning consumption, they try to take the best (optimum) consumption in light of many given facts, including beliefs.
- 2. Rationality of Belief:** When agents adopt decisions concerning beliefs, they try to adopt the best (optimum) beliefs in light of given information.
- 3. Rationality of Information:** When agents make investment decisions concerning information search, they try to make the best (optimum) investment in search given the expected benefits and costs of search.

The isomorphic core of the three layers is a “rational decision.” The rational decision, in each layer, consists of maximizing an objective function, given the inputs (means or constraints).

Table 1 The isomorphic core of a rational decision

Rationality	Means (inputs) →	End (optimum decision)
Rationality of consumption:	<i>Information + Resources + Beliefs</i> →	<i>Optimum Consumption</i>
Rationality of Belief:	<i>Information + Resources + Conceptions/Convictions</i> →	<i>Optimum Belief</i>
Rationality of Information:	<i>Cost/Benefit of Search + Resources + Conceptions/Convictions</i> →	<i>Optimum Information</i>

Table 1 summarizes the core. The rationality of information (first layer) is about deciding what is the optimum information, which becomes an input in the second layer. The rationality of belief (second layer) is about deciding what is the optimum belief, which becomes an input in the third layer. The rationality of consumption (third layer) is about deciding what optimum consumption is.

In all three layers, the direction of the arrow from left-to-right, i.e., the inputs determine the output, i.e., what the optimum decision is. The direction of the arrow entails honesty and integrity: the decision (end) is not allowed to manufacture the facts (means). This definition of rationality implies the separation of means and ends, which others have noted. For example, Gilboa (2010) expressly distinguishes between what is desirable (wishes) and what is feasible (facts) as the core of rational decision making.

The left-to-right direction of the arrow is the case even when one kind of input, in the second layer, consists of beliefs, called here “conceptions,” that are not the product of evidence, i.e., not the product of the rationality of information. Unlike perceptions that are justified by empirical evidence, conceptions are rather directives or rules that the brain produces to assist its cognitive processes. Examples of conceptions include visionary beliefs, cognitive maps, ethical principles, or ideologies.

Conceptions, when infused with internal motivation, can provide the agent with “wishes,” “aspirations,” or in short “convictions” about the potentialities of certain life styles or programs of action (Khalil 2010). Convictions, similar to conceptions, can play the role of input in decision making.

Philosophers of science, such as Thomas Kuhn (1970), identify conceptions as “paradigms” or “matrices.” Ludwig Wittgenstein (1968) identifies convictions as “linguistic games” that individuals use to assist them to gain power and attain their aspirations. While conceptions afford outlines of what is normal science, convictions afford narratives that assist one to seek, aspire, and pursue his or her wishes. While conceptions and convictions lack empirical content, they are imperative in organizing the empirical content. The only way one can defend a particular conception or conviction is not through direct evidence, because there is none. One can only appeal to the efficiency or ease of one’s paradigm or conviction in organizing the available data relative to competing alternative non-empirically based beliefs.²

This is not the place to discuss the philosophy of science on how conceptions differ from data-based beliefs, i.e., perceptions (Khalil 2010, 2013a), or discuss the psychology of internal motivation. It is sufficient to state that social scientists tend to regard non-empirically based beliefs, such as ideology, as beliefs that distort reality (Bénabou 2008). This need not be the case. As long as, following Table 1, the non-empirically based beliefs are restricted to the role of input, they would not be distorting beliefs. Such beliefs would become distorting beliefs if they start to act as products and were allowed to manipulate the data, i.e., reverse the direction of the arrow in Table 1. Then, conceptions and convictions would become misleaders employed in the machinery of self-deception. So, paradigms and even ideologies, as long as they play the role of input, need not be synonymous with self-deception.

4. THREE KINDS OF SELF-DECEPTION

Given the definition of rationality as the left-to-right arrow of deriving the output from input, we can now define what a non-rational choice is, which is the first structural choice. A non-rational choice implies that, first, the rational decision can be identified and, second, the actual action is a negation of such a decision. There can be many reasons behind the failure to enact what is identified as the rational decision. This includes weakness of will, which has been captured in the literature under a quasi-hyperbolic discounting and dynamic inconsistency.

There are different theories of weakness of will or dynamic inconsistency (e.g., Fudenberg and Levine 2006; Khalil 2015b, 2016). But the phenomenon is clear: the agent, under the enormous sensual stimulation of the senses, may undertake a choice that is contrary to the rational decision.

Now, we can define self-deception more precisely. In the first structural choice, the agent, chooses a product other than what is recommended by the *left-to-right* arrow, i.e., the rational decision. As a result, the agent experiences the ill-feeling of self-blame, even before negating the rational decision. In the second structural choice, a deceiving agent, to sustain symbolic utility and nullify the ill-feeling, invents a misleader. The misleader is supposed to manipulate the facts, or provide unsubstantiated interpretation, in order to make the suboptimal choice appear to the conscious self as optimal.

² The role of beliefs as input, and when they should only remain input, is an important question. However, it is a separate question from how to distinguish scientific beliefs from everyday, practical beliefs. The scientific/practical distinction is crucial for identifying the shortcoming of John Dewey’s pragmatism in philosophy and Herbert Simon’s notion of procedural rationality and satisficing in the social sciences (Khalil 2013b).

So, the second structural choice amounts to reversing the arrow. It now runs from *right-to-left*. This means that the agent first chooses the non-rational action and, second, changes the facts *ex post* in order to deceive the conscious self into believing that he or she has chosen the optimal action.

This core definition of self-deception underpins the three layers of decision making. It generates three layers of self-deception, as Table 2 illustrates.

Table 2 The isomorphic core of self-deception

<i>Self-Deception</i>	<i>Means (facts)</i>	\leftarrow <i>End (suboptimum decision)</i>
1. Red herring (reversal of rationality of consumption):	Information + Resources + Beliefs	\leftarrow Suboptimum Consumption
2. Self-Rationalization (reversal of rationality of belief):	Information + Resources + Conceptions/Convictions	\leftarrow Suboptimum Belief
3. Denial (reversal of rationality of information):	Cost/Benefit of Search + Resources + Conceptions/Convictions	\leftarrow Suboptimum Information

In all three layers, we have an agent, wearing the weightless mask, succumbs to a tempting, but suboptimal end. The agent inserts a misleader that allows him or her to manipulate the set of inputs to make the favored but suboptimal end appear as optimal to the conscious self:

- 1. Red herring:** The arrow of causation concerning the first layer, the rationality of consumption, is reversed to allow the suboptimal consumption to manipulate beliefs.
- 2. Self-Rationalization:** The arrow of causation concerning the second layer, the rationality of belief, is reversed to allow the suboptimal belief to manipulate information.
- 3. Denial:** The arrow of causation concerning the third layer, the rationality of information, is reversed to allow the suboptimal information to manipulate the cost or benefit of additional information.

For an agent to commit one kind of self-deception does not entail that he or she must also commit any of the other two kinds.

4.1. Red herring

The most illustrative example of a red herring is the fable of the fox and sour grapes, with which this paper starts. For Gilboa (2010), the fable defines non-rationality. As shown here, the fable, and self-deception in general, is doubly non-rational.

A less commonly noticed illustration of a red herring is the misplaced resentment of the oppressed and victims of injustice. One, e.g., finds him or herself in a situation where an oppressive group abuses members of another group. In the case that it is costly to oppose abuse, one may choose to live with the injustice. But to live with the injustice may, if one is a righteous-minded person, engender self-blame. To suppress self-blame, one would start to blame the victim and even develop hatred or racist attitude towards the oppressed. With such hatred or blaming the victim, one is implicitly using a notion of some fair world, where the oppressed are not actually oppressed, but rather are receiving what they deserve.

Such self-deception springs from what is today called the “just world hypothesis.” The term denotes the observation that people, because of their belief in a just world, tend to rate people, who were “rewarded” randomly in an experiment, as ones who must have higher skills than the rest in the experiment (Lerner 1980). Lerner was surprised to find out that his students derogate the poor as “undeserving.”

Another illustration of the “just world” hypothesis, is when someone cheats another person. In a just world, one cannot cheat. To rationalize it, one may *ex post* invent a new “fact” such as the other person must have cheated first or was just about to cheat.

4.2. Self-rationalization

The best illustration of self-rationalization is the response of Adam when God confronted him for violating his command. Adam responded: “Eve told me it is fine to eat from the tree.” That is, Adam created new erroneous information. Such information is supposed to justify his belief that it is fine to eat from the prohibited tree.

Adam, with such self-rationalization, was reaching for an “excuse” that absolves him from responsibility. In most cases, as in this one, the excuse is usually worse than the suboptimal deed. Adam implies that the information of Eve is superior to God’s. This goes to prove that the agent, if he or she is aware of self-rationalization, would rather opt to take responsibility and suffer from self-blame.

4.3. Denial

With respect to *denial*, the agent changes the cost or benefit of additional information in an *ad hoc* manner in order to justify an already chosen/favored amount of information. Here, the agent allows the suboptimal given information to determine what is the cost or benefit of information.

An example of denial is what Jonathan Baron (2008, pp. 219-220) calls “selective exposure” — as when conservative people mostly read conservative newspapers and liberal people mostly read liberal newspapers. People do not want to know the facts marshalled by the other side. Another example is when people suffering from pain avoid visiting the doctor.

The phenomenon of denial has garnered great attention from Freudian psychologists. They reason that the denial of facts is a “defense mechanism,” where the ego suppresses facts that may arouse shame or anxiety (see Freud 1966).

5. CONCLUSION

This paper argues that self-deception involves two structural choices. First, the agent must have committed a sub-rational act, which invites self-blame. Second, the agent invents a misleader to make the act look as if it were rational and, hence, avoid self-blame. The quest for a blameless choice makes it possible for the deceived person to succeed in hiding the weightless mask, which helps the conscious self to enjoy self-regard and other symbolic utility.

This solves the Self-Deception Paradox. Namely, why people deceive themselves — which makes no sense since self-deception ultimately undermines wellbeing. They are involved in self-deception for the same reason they succumb to temptation and, hence, undermine their wellbeing. With self-deception, the deceiving agent goes one extra step. It

tries to void the negative symbolic utility, given that the injury resulting from succumbing to temptations is already done.

This paper discusses one payoff of the proposed two-structural choice theory. It distinguishes among three kinds of self-deception: red herring, self-rationalization, and denial.

REFERENCES

- Akerlof, G.A. and W.T. Dickens. "The Economic Consequences of Cognitive Dissonance." *American Economic Review* 72 (1982): 307–319.
- Baron, J. *Thinking and Deciding*, 4th ed. Cambridge: Cambridge University Press, 2008.
- Bénabou, R. "Ideology." *Journal of the European Economic Association* 6 (2008): 321–352.
- Davidson, D. "Deception and Division." In *The Multiple Self*, edited by Jon Elster. Cambridge: Cambridge University Press, 1986.
- Davidson, D. "Who is Fooled?" In *Self-Deception and Paradoxes of Rationality*, edited by J. P. Dupuy. Stanford: CSLI Publications, 1998.
- Dupuy, J.-P., ed. *Self-Deception and Paradoxes of Rationality*. Stanford, CA: CSLI Publications, 1998.
- Elster, Jon. *The Multiple Self*. Cambridge: Cambridge University Press, 1986.
- Festinger, L. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press, 1957.
- Fingarette, H. *Self-deception, with a New Chapter*. Berkeley, CA: University of California Press, 2000 (1969).
- Freud, A. *The Ego and the Mechanisms of Defense*, trans. by Cecil Baines, revised edition. New York: International Universities Press, 1966.
- Fudenberg, D. and D. K. Levine. "A Dual-self Model of Impulse Control." *American Economic Review* 96 (2006): 1449–1476.
- Gilboa, I. *Rational Choice*. Cambridge, MA: MIT Press, 2010.
- Hefferman, M. *Willful Blindness: Why We Ignore the Obvious at Our Peril*. New York: Walker, 2011.
- Khalil, E. L. "Symbolic Products: Prestige, Pride and Identity Goods." *Theory and Decision* 49 (2000): 53–77.
- Khalil, E. L. "The Bayesian Fallacy: Distinguishing Internal Motivations and Religious Beliefs from Other Beliefs." *Journal of Economic Behavior and Organization* 75 (2010): 268–280.
- Khalil, E. L. "Two Kinds of Theory-laden Cognitive Processes: Distinguishing Intransigence from Dogmatism." *Behavioral and Brain Sciences* 36 (2013a): 38–39.
- Khalil, E. L. "Practical Beliefs vs. Scientific Beliefs: Two Kinds of Maximization." *Theory and Decision* 74:1 (2013b): 107–126.
- Khalil, E. L. "Temptations as Impulsivity: How far are Regret and the Allais Paradox from Shoplifting?" *Economic Modelling* 51 (2015): 551–559.
- Khalil, E. L. "Weakness of Will and Stiffness of Will: How far are Shirking, Slackening, Favoritism, Spoiling of Children, and Pornography from Obsessive-compulsive Behavior?" In *Handbook of Behavioral Economics and Smart Decision-making: Rational Decision-making Within the Bounds of Reason*, editing by Morris Altman. Cheltenham, UK: Edward Elgar, 2016.
- Khalil, E. L. "Making Sense of Self-deception: Distinguishing Self-deception from Delusion, Moral Licensing, Cognitive Dissonance and other Self-distortions." A mimeo, 2017.
- Konow, J. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90 (2000): 1072–1091.
- Kuhn, T. S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1970.
- Kurzban, R. and C. A. Aktipis. "Modularity and the Social Mind: Are Psychologists too Selfish?" *Personality and Social Psychology Review* 11 (2007): 131–149.
- Lerner, M. J. *The Belief in a Just World: A Fundamental Delusion*. Plenum: New York, 1980.
- McLaughlin, B. P. "On the Very Possibility of Self-deception." In *Self and Deception: A Cross-cultural Philosophical Enquiry*, edited by R. T. Ames and W. Dissanayake, 11–25. Albany, NY: State University of New York Press, 1996.
- Pears, D. *Motivated Irrationality*. Oxford: Clarendon Press, 1984.
- Rabin, M. "Cognitive Dissonance and Social Change." *Journal of Economic Behavior and Organization* 23 (1994): 177–194.
- Rorty, A. O. "User-friendly Self-deception: A Traveler's Manual." In *Self and Deception: A Cross-cultural Philosophical Enquiry*, edited by R.T. Ames and W. Dissanayake. Albany, NY: State University of New York Press, 1996.

- Sartre, J. -P. *Existentialism and Humanism*, trans. Philip Marlet. London: Mathuen, 1948.
- Taslitz, A. E. "Willfully Blinded: On Date Rape and Self-deception." *Harvard Journal of Law and Gender* 28 (2005): 381–446.
- von Hippel, W. and R. Trivers. "The Evolution and Psychology of Self-deception." *Behavioral and Brain Sciences* 34 (2001): 1–16.
- Wittgenstein, L. *Philosophical Investigations*, trans. G.E.M. Anscombe. Oxford: Blackwell, 1968 (1953).

LAKA MASKA SAMOZAVARAVANJA

Samozavaravanje prožima ponašanje osoba, organizacija i država. Samozavaravanje može se porediti sa nošenjem lake maske: dok sama osoba nije svesna maske, posmatrač je obično umeju prepoznati. Čim osoba koja nosi masku postane svesna nje, samozavaravanje nestaje. To dovodi do sledećeg paradoksa: Kako je uopšte moguće da osoba sakrije tu masku od samoga sebe? Da bi se ovaj paradoks razrešio, u ovom radu predlaže se da proces samozavaravanja sadrži "dva strukturalna izbora": 1) agent identifikuje najbolju moguću odluku — ali agent možda neće uspeti (usled uticaja iskušenja) da je odabere, što dovodi do osećanja krivice; 2) agent koji se posmatra može se opredeliti da izmisli neku vrstu distraktora (laž) kako bi ispalo da je izbor optimalan, i time izbegao osećanje krivice. Potraga za izborom koji ne dovodi do osećanja krivice omogućava osobi da uspe i sakrije laku masku od svoje svesti. Pored rešavanja Paradoksa samozavaravanja, predložena teorija o dva strukturalna izbora daje uvid u tri vrste samozavaravanja: diverziju, racionalizaciju, i poricanje.

Ključne reči: *racionalni izbor, samooptuživanje, iskušenje, upravljanje slikom o sebi, hipoteza o pravičnom svetu.*