Keeping Secrets from Ourselves: Understanding Self-deception Through Theory, Evidence and Application

Mathew J. Creighton

Abstract
Self-deception is a difficult concept to share with students. Although few students find it implausible that they are capable of keeping secrets from themselves, the social theory, application, and practical demonstration of self-deception is far from straightforward. This work offers a three-step approach to teach a theoretically-grounded, evidence-based, and application-reinforced understanding of self-deception. Rooted in work on identity by Mead (1934), the approach outlined here engages with interdisciplinary case studies derived from social psychology (Greenwald, McGhee and Schwartz 1998) and behavioral economics (Ariely 2012). The theory and case studies build toward a peer evaluation that offers students a concrete demonstration of self-deception with implications at the individual- and group-level.

Keywords
behavioral economics, deception, lies, lying, G. H. Mead, metacognition, pedagogy, peer evaluation, secrecy, secrets, self-deception, social psychology

Self-deception presents a pedagogical problem. Few students struggle with the idea that they might be able to deceive themselves. But to move from intuition to a grounded understanding of self-deception requires that students reconcile their perception of reality with some difficult (and occasionally uncomfortable) tenets. Most notably, self-deception necessitates an acceptance there are secrets kept by ourselves from ourselves. Is this even possible? How
can one initiate an internal “cover-up” if one is unaware of what exactly is being covered?

Academic approaches to lying and deception generally focus on cases that involve more than one participant (i.e., lying to others). An exception in sociology would be Simmel (1906), who alludes to self-deception as a form of illusion,¹ suggesting the presence of some sort of false interpretation of reality. Later and oft-cited work by Bok (1999) offers an example of a succinct definition, stating that lying is an intentionally deceptive message in the form of a *statement*. More recent work by Carson (2010) quibbles little with Bok’s (1999) premise, but establishes the importance of the *audience* by necessitating an expectation that someone needs to believe the statement to be true. However, in both cases do not account for self-deception. One would need to speak/write to oneself and the ability to delineate the source and recipient (i.e., audience) necessitates some form of cognitive dissidence. Who would provide the *statement* or the *audience*? Rather than an observable event, requiring action and interpretation, self-deception is a beguilingly internal process.

Clearly, self-deception is a bit more complicated than general approaches to define secrets, lies, and deceits. That said, self-deception is increasingly at the center of weightier topics in social science, such as the unconscious perpetuation of racial, ethnic and gendered marginalization. Providing a theoretical basis and empirical framework offers the crucial foundation upon which students can build an understanding of the implications of unconscious biases, the premise of

---

¹ Funkhouser (2019) embraces this tradition, referring to self-deception as *self-delusion*. 
motivated reasoning, and the fine line between intentional and unintentional deceit. This work draws from my teaching experience to offer applications, case studies, and theoretical frameworks that have proven useful in exploring self-deception. However, to suggest that this is the only way to approach the subject would, in of itself, indulge a certain degree of self-deception.

**Theoretical Foundation: Self-deception**

Eric Funkhouser (2019), the esteemed philosopher, and the character of Vizinni, played by Wallace Shaw in the film *The Princess Bride*, might not be the most intuitive pairing. However, both agree that self-deception is rooted in the inability of an individual to conceive of some underlying truth. Vizinni, when confronted with information that contradicts his a priori assumptions reacts by stating “inconceivable.” The insight (and the humor) results from befuddlement expressed by his traveling companions upon being told something is inconceivable. Their demeanor changes suddenly and visibly as they wipe new information from their mind and shift their perception to share the self-deceived version Vizinni considers conceivable. Funkhouser (2019) sees this process as more of a “failure of self-knowledge,” underlining the fact that identifying self-deception reveals gaps between objective reality as it relates to the individual in question, and that same individual’s perception/articulation. Examples are wide ranging, including disconnected notions of physical or mental ability (e.g., I am a stable genius) and some plausibly objective measure (e.g., consistent behavior). There are implications beyond the individual as self-deception is a component of
complex social phenomena linked to core areas of social and sociological enquiry - race, gender and class.

The work of G.H. Mead (1934), which resides at the intersection of sociology, social psychology, and philosophy, provides a useful theoretical framework.\(^2\) The premise is a layered understanding of identity, which is typically visualized as concentric circles. At the center resides the private “I,” surrounded by the “me,” which is in turn encircled by the “self.” A core contribution of Mead is to our understanding of social organization as a dynamic process rooted in interaction. The outer two layers of identity – the “me” and the “self” – are fundamentally defined by one’s ability to anticipate/inhabit the perspective of others, which, in turn, shapes behavior. In other words, we regulate our conduct, including expression, as a response to the social context within which we find ourselves. This is helpful for understanding many forms of deception, which require a clear statement and an audience, but is less helpful when considering self-deception.

Rather than providing a clear sociological foundation to situate self-deception, the value of Mead’s work is to highlight the difficulty in accounting for aspects of identity that are unknown or unacknowledged by an individual. In other words, social theory finds social deception as difficult as my students. We can intentionally and strategically present ourselves to others and this is the key to managing our identity in the eyes of others. However, our private self - the “I”

\(^2\) Of note, there is a tradition in sociology that emerged from Mead (1924) that morphed into observational, dramaturgical articulation of social interaction (e.g., Goffman 1963) or empirical formalism (e.g., Stryker 1980). Although crucial to provide a broader understanding of the expression of identity, neither tradition in the literature offers additional insight into the role/emergence of self-deception beyond what a close examination of Mead (1934) provides.
– contains our secrets. For these secrets to be self-deceptive we must now delve into the unconscious.

Students can struggle with this aspect of self-deception so I use the odd case of Jered Threatin, who lived in a world in which he was a celebrated musician (Luseenhop 2018). This example depicts a person who, despite accessible evidence to the contrary - the complete absence of fans - sincerely believes in his own fame and success. Threatin booked a road tour during which he was repeatedly surprised by empty venues, always finding a technical or logistical explanation. Students are understandably skeptical of the veracity of Threatin’s self-deceit. That is a reasonable response, but it underlines Funkhouser’s (2019) and Vizinni’s point that self-deception is rather hard to understand in practice. One could almost imagine that Jared Threatin’s replying “inconceivable” when queried about the absence of an audience.

The case of Jared Threatin, because it is intentionally low-stakes, offers a bridge to weightier topics, namely, unconscious bias. A core driver of social and economic inequalities in the US and Europe is systematic marginalization based on race (Reskin 2012; Pager and Shephard 2008), gender (Reskin 1993), and class (Elo 2009). Despite legal intervention, inequities have proven remarkably resilient (Cohen 2013) and attention in the literature has turned to unspoken and/or unconscious pathways by which marginalization can persist (Bonilla-Silva 2009; Burke 2017). The key detail, which diverges from any earlier focus on structural inequalities, is the widely-documented simultaneous denial and existence of biases (Blanton and Jaccard 2008; Bonilla-Silva 2009; Bonilla-Silva
2015; Hartman et al. 2017). Self-deception provides the micro-level mechanism by which macro-level marginalization can emerge and, moreover, perpetuate. It is here that I find a detour into case studies to be useful.

Case Studies: Implicit Associations, Learned Self-deception, and Catching Yourself in a Lie

Self-deception is not easily measured (Bartlett 2017; Blanton and Jaccard 2008). The difficulty arises from the need to show that one has a belief/perspective/attitude of which one is unaware or, alternatively, has become unaware of over time. Self-deception could be a permanent, constant state, or a divergence over time from something that was known. In the latter case, one becomes intentionally or unintentionally removed from one’s consciousness. To avoid an overly abstract description, two cases studies have proven particularly illuminating.

The first case is a stalwart of social psychology termed the implicit association test (IAT). Greenwald, McGhee and Schwartz (1998) introduced the approach, and subsequently Greenwald et al (2015) underlined the implications at the macro-level. The logic is straightforward and the basic design is premised on the idea that we experience a greater cognitive load when linking concepts that we don’t naturally associate. If, for example, we don’t have positive associations with a specific gender, a measurable delay in linking positive words/images emerges. The implications for individual-level identification of self-deception is clear in that one can take the test and have their own differences in
response times measured. If a delay/bias arises, discordance with one’s self-concept would underline whether self-deception is a plausible interpretation.

The extent to which the test is accurate and interpretable is contested (Oswald et al. 2013), as is its use in higher education (Bartlett 2017). Although the IAT might not offer a statistical truth serum, it does raise interesting questions about our ability to compartmentalize our biases. Students find numerous online IATs readily available on a variety of topics, which provides a very concrete link between the learning experience and more abstract efforts to measure self-deception. Admittedly, many remain skeptical of the value of the IAT and some outright refuse to engage with it out of a sense that uncomfortable biases might be revealed.

A second case that evidences self-deception is provided by Ariely (2012). Rather than the IAT, which operates as a sort of infallible indicator of latent tendencies, Ariely (2012) draws from experimental economics to map out the emergence of self-deception. The approach is straightforward. Students take a simple, timed test. Randomly, some are given the answers and asked not to use them until after the assessment. Some are provided no answers. Predictably, those with access to answers during the assessment do better on average, indicating that the temptation to cheat (and lie about it) was more than some could resist. The key insight into self-deception comes in the next step.

The same students are subsequently given a similar assessment, but no answers are provided. The absence of any answers is made clear. In addition, the students are asked to evaluate their anticipated performance. Interestingly, the
this same group of students who did better due to cheating now report significantly higher average expectations. Although no actual differences, on average, emerge in the follow-up test, it seems that the students who deceived others by cheating in the first test had come to believe their own lie. Students now report expectations akin to what Funkhouser (2019) referred to as a “failure of self-knowledge.”

In the end, these two cases offer students concrete cases of plausible self-deception. IAT provides an example, albeit contested, of a direct measure of unconscious attitudes. The approach is widely used and can even be attempted in a classroom setting given the multitude of online opportunities available. Ariely (2012) provides an instance of the emergence of self-deception over time. Rather than offering evidence of an unconscious repository of biases, Ariely’s (2012) approach points to a more mundane evolution in which deception morphs into self-deception. Both cases ground a largely theoretical and abstract depiction of self-deception, laying the foundation for the next step - an application of the concept.

**An Application: Self-deception and Peer Evaluation**

The study of self-deception provides insight to the students about how their own self-deception might impact others. The strategy pursued here looks to link individual-level behavior with group-level patterns. To do so, a larger class is helpful as size provides more students and the option for randomization into groups. The latter is particularly important to avoid, on average, the need to
account for bias attributable to pre-existing social ties between group members. That said, peer evaluation should be used with caution.

In practice the implementation of peer assessments are difficult in large-format teaching.\(^3\) For teaching self-deception, the approach has proven very useful in re-enforcing and deepening student understanding of the self-deception as randomization and group-level comparisons necessitates as large number of participants. The design used is shown in Figure 1 and described below.

**Overview of Preparation and Delivery**

All enrolled students in a 3rd-year undergraduate module in sociology are randomly assigned to groups of between four and five students. The randomization was implemented via the virtual learning environment (or VLE; e.g., Blackboard). The use of randomization is crucial as self-selection into groups would limit the utility of peer evaluation as dynamics unrelated to the group-level interaction would plausibly influence/domininate the evaluation (i.e., friends would reward friendship).

---

\(^3\) See Mostert and Snowball (2013) for a recent review.
Each of the groups is assigned a four-part group project that requires regular interaction and collaboration/delegation to complete. In addition, the project entails a single mark that is given to each participating member. Students are informed that a peer evaluation will accompany the group mark at the onset of the project.

Each group submits a final project, which is marked and returned to the students with each of the four component parts being assessed and specific feedback offered (i.e., one to three paragraphs of comments per component part). Subsequently, all group members complete a peer evaluation form implemented via WebPa, which is a VLE extension that generates interactive
forms. The full roll-out was preceded by a single “pilot” in which a student completed the assessment under supervision to assess interpretation and user interface. Appendix 1 shows the text of the peer evaluation.

The final step in the exercise is the use of self-evaluation, which is key for highlighting self-deception. Specifically, students are asked to give themselves a mark for how they anticipate other group members will rate their overall performance. This occurs sequentially after marking their peers and is based on a combination of contribution and responsiveness – as well as any additional aspects of the team-based collaboration that these two categories neglect. The meaning and interpretation of contribution and responsiveness are reinforced in lecture prior to the peer evaluation.

Reflection on the Peer Evaluation

Once the responses to the peer evaluation are submitted, the next step is to reflect on the results. All evaluation is anonymous, which is known to be the case by all students, as only the lecturer is privy to individual-level marks. The marks reported to students are the average across all group members, which prevents any single group member’s mark from being identifiable. The results offer two dimensions of interpretation.

First, students are provided the opportunity to ponder their work from the perspective of others in their group. This enters the theoretical space of Mead (1934) in that the “self” is being brought into focus, which requires the acknowledgement of the role of context and interaction. Second, the results offer
a crucial demonstration of self-deception as systematic patterns at the group-level are observable. One outcome to highlight pertains to gender, but other details are plausibly useful so long as they form part of the data collected in the evaluation (e.g., psychosocial measures, attitudinal measures). The motives and emergence of self-deception would necessarily be sensitive to context, but three patterns could potentially emerge from the exercise.

First, students could have a good sense of how their peers see their contribution and responsiveness, indicated by a close match between their anticipated and the actual peer evaluation. Alternatively, students could envision their work receiving a mark that, in the end, is higher than that which their peers give them. This would be a positive form of self-deception. The third option is a similar mismatch, but negative in that a student could foresee receiving a more negative mark than that which results from the peer evaluation. Of course, at the individual-level, one could argue that in any given specific case it is the peers who are deceptive in their assessment. However, at the group-level, trends are not so easily explained.

For example, results for some cohorts demonstrate that male students report an average an expectation that their work will receive a higher mark than they subsequently receive from their peers – male or female. In some cases, female group members report a lower average expected mark than they actually receive. In short, although context would undoubtedly shape this pattern, male students are positively self-deception and female students report a more negative views, on average. Of course, this is not a definitive example nor does it offer a water-
tight link between individual-level self-deception and systematic biases at the group-level. It is, however, suggestive and has proved a useful exercise to re-enforce student understanding of self-deception rooted in theoretical and case studies from the literature.

The peer-evaluation application offers insight into a pattern very likely to be shaped, at least in part, by unconscious biases. This delivers insight into a key aspect of how self-deception bridges the micro and macro. The example given of differences in self-deception by gender plausibly reveals a pattern emerging from something other than the experience of the group project - something about which we are unaware. In addition, the peer-evaluation exercise and post-evaluation of results surpass the utility of online (e.g., IAT) or taught case studies in that some students experience the implications of their own self-deception for others. In my experience, there is rarely a substitute for application in the learning process.

**Conclusion: An Outline for Teaching Self-deception**

Self-deception is a concept that resides at the intersection of philosophy, sociology, behavioral economics, and social psychology. At the onset, there is an intuitive dimension that emerges from widely-used every-day experiences often linked to overly optimistic self-evaluation (e.g., so-and-so thinks they are good at [insert skill], but they are really [insert some plausibly objective assessment]). However, when the formality of defining self-deception comes to the fore, intuition gives way to some very tricky philosophical questions. Funkhouser
(2019) and Vizinni help, but it is not easy for students to grasp how one can be unaware of a belief. This lack of self-awareness erodes the foundation of conviction, that is, our belief in our grasp of objective reality. Overcoming this difficulty in understanding is crucial as self-deception provides the key ingredient to a robust grasp of widely-used, hard-to-define concepts such as unconscious bias and color-blind racism.

The approach outlined here offers an engaging pathway that can concretize student understanding of self-deception, defined by three steps. First, social theory offers the foundation. In particular, the layered perspective on identity defined by the work of Mead (1934) provides framework that distinguishes private and internal perceptions of self (i.e., “I”) from those shaped partially or mostly by interaction with others (i.e., “me”, “self”). Granted, Mead (1934) is a shaky foundation as the “I” is not clearly capable of providing shelter for beliefs/perceptions untouched (or untouchable) by self-awareness. That said, the theoretical perspective provides a structure and orients students toward the difficulty of conceptualizing self-deception.

The next step involves the introduction of the unconscious, which benefits from methods-driven case studies. Two cases have proven particularly useful. The first is the implicit association tests (IAT), which offers a mainstream claim of achieving a behavioral truth-serum. Students can easily understand the method and quickly complete an IAT online, but skepticism in the method (among students and in the literature) necessitates an additional case study. Behavioral economics (e.g., Ariely 2012) offer a somewhat more straightforward approach as
it involves student assessment and the emergence of patterns in self-deception based on an observable conditioning process. Self-deception is shown to be a dynamic process, conditioned via the performance of acts of deception over time.

The third and final step involves an application via a peer-evaluation exercise. The goal is to offer direct evidence (albeit imperfect) of systematic biases in self-perception from which students are plausibly unaware. Moreover, the peer evaluation allows students to experience how their own biases plausibly affect others.

Perhaps it is an exercise in self-deception to adequately teach self-deception in a few short lectures. That said, I have found the effort worthwhile. There are obvious imperfections in the approach outlined here, but it has resulted in a number of cohorts of students who are capable of grasping and understanding the wider implications of self-deception in their own life and in the lives of others. This is distinct from a standard overview of deception/lying that involves deceiving others and, by definition, necessitates reflection. Self-deception as a topic has proved to be engaging and hopefully the more light-hearted examples (e.g., Jared Threatin) can offer engagement without dominating the topic or de-emphasizing more weighty societal problems (e.g., systematic marginalization based on race, gender, and class). Appendix 2 lists the materials and their sequence of introduction in the forma of an example syllabus.
References


https://www.chronicle.com/article/Can-We-Really-Measure-Implicit/238807


Appendix 1.
Text of Peer Assessment

What mark is most appropriate for the RESPONSIVENESS of each group member? RESPONSIVENESS refers to communicating with other groups members and meeting when needed. [This directly counts toward the peer assessment]
1. E
2. D
3. C
4. B
5. A

What mark is most appropriate for the CONTRIBUTION of each group member? CONTRIBUTION refers to doing work that is high quality and equal to other group members. [This directly counts toward the peer assessment]
1. E
2. D
3. C
4. B
5. A

What mark is most appropriate for OVERALL for each group member? OVERALL refers to responsiveness, contribution and any additional aspects of the project not covered by those two categories. [This directly counts toward the peer assessment]
1. E
2. D
3. C
4. B
5. A

From your perspective, what mark will each group member give you OVERALL? OVERALL refers to responsiveness, contribution and any additional aspects of the project not covered by those two categories. [This is used for reference]
1. E
2. D
3. C
4. B
5. A
Appendix 2.
Example of Syllabus Outline for Theory and Case Studies for Teaching of Self-deception

Theoretical foundation: Self-deception

Lecture 1

Mead (1934), 164-186.
Simmel, (1906).

Lecture 2

Goffman (1964), 1-40; 91-104.
Stryker (1980), 51-84.

Case studies I: Implicit associations, learned self-deception and catching yourself in a lie

Lecture 1

Lussenhop (2018).
Ariely (2012), 141-163.

Lecture 2

Bok (1989), 59-73.
Carson (2010).

Supplemental

Case studies II: Implicit associations, learned self-deception and catching yourself in a lie

Lecture 1

Supplemental
Lecture 2

Bartlett (2017).
Oswald et al. (2013).