

Manipulation and Deception

Shlomo Cohen

To cite this article: Shlomo Cohen (2018) Manipulation and Deception, Australasian Journal of Philosophy, 96:3, 483-497, DOI: [10.1080/00048402.2017.1386692](https://doi.org/10.1080/00048402.2017.1386692)

To link to this article: <https://doi.org/10.1080/00048402.2017.1386692>



Published online: 15 Oct 2017.



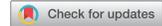
Submit your article to this journal [↗](#)



Article views: 372



View Crossmark data [↗](#)



Manipulation and Deception

Shlomo Cohen

Ben-Gurion University

ABSTRACT

This paper introduces the category of ‘non-deceptive manipulation that causes false beliefs’, analyzes how it narrows the traditional scope of ‘deception’, and draws moral implications.

ARTICLE HISTORY Received 11 July 2016; Revised 31 August 2017

KEYWORDS manipulation; deception; ethics of veracity

1. Introduction

The concept of manipulation remains significantly underexplored in philosophy in general and in moral philosophy in particular.¹ Considering the pervasiveness of manipulations in all spheres of human interaction, this neglect demands rectification.² The understanding and assessment of manipulation, beyond their singular value, can shed valuable light on central aspects of social intercourse (from education and child rearing to negotiating, advertising, democratic process, and more). This paper pursues one morally central angle of this exploration: it examines how reflection on manipulation can illuminate the boundary between manipulation and deception and, through it, the ethical boundaries of misleading communication.

Manipulation has too often been understood as ‘a kind of *deceptive* non-coercive influence’ [Coons and Weber 2014: 10]. That this cannot be true, however, is shown by numerous mundane examples of unconcealed emotional manipulation (at times involving active cooperation by the manipulated), non-deceptive yet manipulative advertising, and the like. Once we acknowledge non-deceptive manipulations [Noggle 1996; Gorin 2014a], we may either view deceptions as a distinct subset of manipulations, or view the two as separate entities.³ In either case, the common tendency is to see the intentional causation of false beliefs as the defining feature of deception, which sets it apart from manipulation. This paper rejects that diagnosis, introduces the category of *non-deceptive* manipulations that cause false beliefs,⁴ and argues that it better

¹ Patrick Todd’s [2013] diagnosis is correct: ‘direct philosophical treatments of the notion of manipulation are few and far between.’

² The first philosophical book on the concept of manipulation is very recent: Coons and Weber [2014].

³ I favour the former understanding, but shall not pursue that debate, as it is inconsequential for any substantial claim in this paper.

⁴ Two terminological clarifications: (1) if we viewed deception and manipulation as separate entities, then ‘non-deceptive manipulation’ would obviously be redundant; (2) hereafter, causing ‘false beliefs’ refers to false-beliefs-that-are-believed-to-be-false.

accounts for very many instances traditionally considered to be deception. Introducing the new category will induce a re-evaluation of the field of deception, suggesting a significant narrowing of its scope. This will simultaneously allow a more complex understanding of the ethics both of manipulation and of deception.

Before commencing, I should emphasize that recognition of the sheer existence of manipulations that cause false beliefs is independent of possible disagreements on how precisely to distinguish manipulation from deception. Consider this: Paul intends to manipulate Mary emotionally (for example, into liking Paul). Paul's actions cause Mary to develop certain false beliefs, although this was no part of Paul's intention. Lacking that intention, his action is not deception;⁵ yet it is (intentional) manipulation that causes false beliefs. Manipulations that cause false beliefs are clearly not *ipso facto* deceptions.

2. Non-Deceptive Manipulations that Cause False Beliefs, and Non-Propositional Communication

Let us start with two intuitive characterizations of deception:

A: Deception is the intentional act of convincing another of a falsehood.

B: Deception is the intentional causation of false beliefs in another.

While, as I will argue, *A* is the more precise formulation, *B* is the one that regularly appears in definitions of deception.⁶ Examination of what *B* entails will show that definitions of deception based on the idea of 'intentionally causing false beliefs' conceal a significant ambiguity between convincing someone of a falsehood and merely manipulating someone into holding a false belief, without a similar dynamic of convincing. This distinction has not been clearly recognized in accounts of deception. Simple examples will show that the latter characterization misses essential components of deception, and is therefore over-inclusive. Traditional definitions of deception therefore refer to a heterogeneous group that includes deceptions alongside non-deceptive manipulations. Sifting manipulations out of the definition of deception, by basing the definition on the essential component of intentionally convincing someone of a falsehood (that is, on *A*), will better capture the essence of deception.⁷ Accordingly, the general course of the argument of the first three sections will be to start with *B* and, through a series of examples and arguments, will narrow it to *A*.

The distinction suggested is ethically significant. As I will discuss, deception as such carries heavier moral weight compared with manipulation: normally deception is morally worse than manipulation, and quite often manipulation isn't even impermissible.⁸

⁵ I follow the vast majority of thinkers in viewing intention as necessary for deception.

⁶ Hence, for instance, Thomas Carson [2010: 50] defines deception thus: 'A person *S* deceives another person *S1* if, and only if, *S* intentionally causes *S1* to believe *x* (or persist in believing *x*), where *x* is false and *S* does not believe that *x* is true'. See also James Mahon's [2007] review of definitions of deception.

⁷ I will forego providing my own definition of deception; this would require discussing subtle points wholly tangential to my focus. Rather, no matter what the precise definition, recognizing the category of non-deceptive manipulations that cause false beliefs will require adding clauses that restrict the scope of the traditional definitions, in ways discussed below.

⁸ Although *manipulativeness* is admittedly a vice, this should be understood as characterizing not the person who manipulates (*simpliciter*), but instead the person who manipulates improperly: much manipulation is morally unproblematic, as discussed below.

Given the tremendous ubiquity of manipulative communication in all aspects of human interaction, the fact that extant discussions of deception leave this ethical distinction below the radar (inevitably so, since the conceptual distinction is ignored) is deeply flawed and morally consequential, as my analysis will make clear.

Definitions of deception based on its characterization as ‘intentionally causing false beliefs’ almost invariably are over-inclusive, in the following sense: they allow the possibility of causing false beliefs while ‘bypassing’ or ‘not involving’ the agency of the victim (for example, via brain stimulation).⁹ But causing false beliefs by directly manipulating the neurological hardware that supports belief-formation does not fall under the conventional meaning of ‘deception’; rather, deception necessarily happens through some communicative process that engages the victim’s power of judgment. The distinction between engaging and not engaging agency (judgment) is valid and important, only it is very coarse: many cases do not conform to the dichotomy that it imposes. We can cause false beliefs intentionally without bypassing the victim’s agency—indeed, while engaging in interpersonal communication—and this would still *not* qualify as deception. The condition of not bypassing agency is necessary but not sufficient for deception. This condition indeed does sift out some manipulations that cause false beliefs—for example, those of the ‘mechanical’ handling of the brain, but not others, those that comprise the vast majority. Further distinctions about *how* agency is engaged are pertinent, and must be clarified.

Consider the following case:

Pill. Cognizant of people’s tendency to associate the colour blue (more than, say, the colour orange) with tranquillity [De Craen et al. 1996], pharmaceutical company X manufactures blue tranquilizer pills.

Predictably, marketing blue tranquilizer pills causes the public to buy more of them than the rival company’s orange pills—coming to view them, falsely, as more potent. Intentionally invoking the false belief is achieved here without transgressing truthfulness, as, simply put, there are neither ‘truthful’ nor ‘untruthful’ colours to pills. Pills must have *some* colour. X chooses blue, and this is perfectly legitimate. The shrewd X invokes a form of cognitive-perceptual bias, biasing people’s judgment into believing falsely that X’s pills are more potent than the rival’s identical agent. Since nothing in marketing blue pills deviates in any way from standards of veracity, there is no deception. And yet judgment was surely manipulated. (Again, it was *not bypassed*, as in creating false beliefs through neurosurgery, pharmacotherapy, etc.) *Pill*, therefore, is a case of non-deceptive manipulation that causes false beliefs.

(But, granted that there is no breach of veracity in expression, isn’t there nonetheless deception in concealing the grounds for the choice of colour? There is not. The initial plausibility of this reservation rests on the idea that some deceptions are not (only) about what is said but are about the state of mind of the deceiver. But whereas the relevant state of mind in those cases is one that refers to the truth of what is stated—that is, ignorance of it, or lack of concern for it (bullshitting)—and this is indeed pertinent to the warrant of truth and hence to deception, here, in contrast, such knowledge of or concern for the truth is irrelevant, as there is no truth of the matter (see further elaboration below). The relevant state of mind in our example is, rather, X’s *motivation* regarding the best choice of colour; and such transparency of motivations is emphatically not

⁹ See Mahon [2007: 185, 192]. Carson’s definition above is a good example.

integral to truthfulness in communication, although in specific cases it might be—again, see further discussion below.)

Consider another example:

Scent. You aspire to sell your home above its market price. Expecting that savvy potential buyers will not evaluate it as highly, but having read about useful marketing techniques, you make your home smell like fresh pine [Herrmann et al. 2013]. It works: potential buyers wrongly estimate your home at above market price.

Had you concealed some defect in the house, or provided wrong answers to inquiries about it, you would have deceived the buyer; but you did no such thing. The buyer truly saw the house for what it is—needless to say, the buyer is free to preserve the scent of fresh pine once she moves in, so that what she gets would be *precisely* what she saw (and smelled). While you did not deceive, you surely manipulated the buyer. *Scent* involves non-deceptive manipulation that causes false beliefs.

My claim that *Scent* and *Pill* involve manipulation was based on intuition and common sense. Let us attempt greater precision. The notion of manipulation is hard to define; it has even been argued that it may be an undefinable ‘family resemblance’ concept [Todd 2013; cf. Coons and Weber 2014]. Perhaps the most widely used characterization goes approximately like this: to manipulate is to influence another’s choices, beliefs, attitudes, etc. in a predictable direction, by means that are neither coercion nor rational persuasion.¹⁰ This characterization is plainly crude (and negative, to boot). Moreover, it is problematic for at least two reasons. (1) Given that judgment as to what should be considered a good (weighty, convincing) reason is influenced by emotional attitudes [Frijda, Manstead, and Bem 2000], and is inescapably shaped by situational and interpersonal causal influences, it is not clear that the concept of *pure* rational persuasion is not an empty set in human matters (that is, outside of logic, mathematics, etc.).¹¹ (2) Rational argument can itself be used manipulatively [Tsai 2014; Gorin 2014b]. It is therefore arguably preferable to understand manipulation as intentionally influencing someone in a predictable direction by means that compromise good judgment.¹² We can subsequently describe the difference between deception and misleading manipulation thus: while the latter compromises good judgment by interfering with its standard functioning and by inducing suboptimal judgment (interfering with its ‘form’),¹³ deceptions undermine judgment by interfering with its input (‘content’): they do not change its standard functioning; rather, they prevent its successful conclusion by feeding it the wrong data. Both kinds of interference make judgment go astray (form false beliefs), but they do this via different mechanisms. This difference in mechanisms, between hindering judgment through its ‘form’ or through its ‘content’, will help to sift manipulations out of the realm of deceptions. This account confirms our earlier point: the more relevant and interesting question about causing false beliefs is not that of whether it engages or bypasses agency (judgment), but rather that of *how* it engages (interferes with) agency.

There is nothing illogical in postulating that all intentional creation of false beliefs is deceptive: ‘non-deceptive manipulation that (intentionally) causes false beliefs’ would

¹⁰ This formulation owes to Faden and Beauchamp [1986: ch. 10].

¹¹ To quote Sarah Buss [2005: 214], ‘No rational chooser can do anything without the aid of nonrational influences.’

¹² This has an affinity with Noggle [1996].

¹³ For example, by deflecting attention from certain considerations or by fixating on others (see examples below).

then be an oxymoron. But we expect our classifications to be consonant with common linguistic sensibilities, and the example of inducing false beliefs by neurobiological means already showed that not all intentional creation of false beliefs falls under ‘deception’. *Pill* and *Scent* introduce a much deeper restriction, as they, too, are intuitively not deceptions. (Relatedly, if they were deceptions, they could not be licit marketing strategies.) *Pill* and *Scent* present manipulations that cause false beliefs. Their legitimacy is not merely conventional; it expresses the fact that they communicate nothing untrue. That their not communicating anything untrue is *essentially* the case relies on the fact that they employ *non-propositional* mechanisms for creating false beliefs. This is an important point that merits elaboration.

‘Linguistic’ deception is deception through language. ‘Non-linguistic’ deception includes all intentional causation of false beliefs through gestures, silence, rearranging the physical space, and so on.¹⁴ All linguistic deception is, trivially, propositional: that is, it affirms or denies something, and can therefore be significantly characterized as delivering falsehoods (when it does). The point to emphasize here is that *non-linguistic* communication, too, can be considered propositional in an extended sense: the propositional content, although not explicit, is based on shared paralinguistic practices and conventions of what, probabilistically, we take specific gestures to mean in given contexts. In non-linguistic deception, false beliefs are created as a reaction to those inferred false propositions. Kant’s famous example of misleading by packing his suitcase in plain view is a paradigm of non-linguistic yet propositional deception: it ‘imitates’ language, in the sense that it is commonly understood (and known to be so understood) that, in standard relevant situations, packing one’s suitcase conveys the propositional meaning ‘I am about to go on a journey.’ Probably, most non-linguistic deceptions are propositional in this (extended) sense. In contrast, *Pill* and *Scent* are *not* propositional: although colour or scent can affect people’s judgment in predictable ways—and can therefore be used to manipulate—the false beliefs created here cannot be similarly traced back to an implicit proposition that was non-linguistically conveyed (that is, in contrast to Kant’s non-linguistic deception, here the fact that perceiving the colour blue triggers connotations of tranquillity surely cannot amount to claiming that the false proposition ‘our pills are more potent than the competitor’s’ was, even implicitly, *expressed* and transmitted as such).¹⁵ Being non-propositional, these cases of manipulative communication have *no truth value*. This undergirds the intuition that they cannot possibly qualify as deceptions.

It might not be surprising that the category of non-deceptive manipulation that causes false beliefs, despite its ubiquity, escaped recognition, given that non-propositional cases of creating false beliefs have not been analysed in the literature on truthfulness, while they alone are categorically not untruthful, as they lack truth value altogether. These manipulations influence others into making judgments (a clear engagement of agency) that are biased and that result in false beliefs, without expressing any message. Not delivering a message (and *a fortiori* not being untruthful), these

¹⁴ In exceptional cases, language itself can be used non-linguistically: performing linguistic acts can be used to *show* something unrelated to the conventional meaning of what is said: see, e.g., Searle [1969: 44–5].

¹⁵ There is a trivial sense in which non-propositional influences can be called propositional: e.g. ‘This house smells like pine’ or ‘This pill is blue.’ But this is unhelpful: unlike the Kant example, *the contents of these (true) propositions are not the contents of the false beliefs created.*

manipulations are not deceptions (just as mechanical manipulations of the brain are not deceptions) even if they cause false beliefs.

To recapitulate: it is correct to limit deception to intentionally creating false beliefs through a communication process, where the victim's agency is engaged by the deceiver. But the creation of false beliefs can be all of this, and still not qualify as deception—it might be manipulation.¹⁶ The following discussion will review the traditional field of deceptions, to determine what else within that field should be similarly reclassified.

3. Non-Deceptive Manipulations that Cause False Beliefs via Communicating Truths

False beliefs can be intentionally created in different ways. Two main parameters participate in drawing an adequate classification: the nature of what is communicated, and the nature of its connection to the false beliefs created. That which is communicated can be either a falsehood (a false proposition) or not, and its connection to the false beliefs created can be either direct or not. I will argue that deception is restricted to cases where a falsehood is communicated *and* it is the direct cause of the false belief; other cases are manipulations. This will constitute an explication of the idea of deception suggested at the outset by characterization *A*: namely, deception as convincing the other of a falsehood. Manipulations that cause false beliefs fail this condition.

As a preliminary to explaining this claim, two definitional clarifications are due regarding the notion of 'false (true) proposition'. These are meant to shift focus from technical distinctions based on form of expression to the essence of what is communicated, the truthfulness of which we want to assess. The first was already presented above: to make good sense of the field of deception, 'proposition' is best understood in an essential sense that includes the aforementioned kind of non-linguistic expressions; 'proposition' should not be understood in the restricted technical sense of making actual statements. Again, this typology is helpful when attempting to focus on the more relevant and interesting fault-lines in the domain of untruthful communication (that is, the question of whether a statement was in fact made is more 'technical' and less illuminating than the question of whether hearers can competently identify an implicit propositional message and its truth value). The same motivation leads to the second distinction, between 'true' and 'false'. The simple suggestion here is that, for purposes of distinguishing deceptions from manipulations, it is best to use '*false* proposition' not in the narrow sense referring exclusively to false assertions but also to what is literally true yet constitutes a false implicature. The false proposition in the latter case will refer to the conventional interpretation of what is said which competent hearers will standardly form in the given context.¹⁷ (This classification will obviously not prevent us

¹⁶ One might doubt that non-propositional influence at all works by forming *beliefs*. In response, (1) granting that non-propositional communication (sometimes) influences non-cognitively, it is sufficient for my argument that *some* cases influence cognitively; (2) there is sound empirical evidence for this dynamic (e.g. Krosnick et al. [1992]); (3) certain authorities even argue that *all* such influence is cognitive (e.g. Mitchell et al. [2009]); (4) various theoretical models explain the structure of the cognitive process (e.g. Bizer et al. [2003]); (5) regardless of such models, it is enough that one predicts correctly that, say, an unwarranted negative thought will more likely be primed consequent to introducing noxious stimuli, and that one intends this effect, for there to be non-deceptive manipulation that causes false beliefs.

¹⁷ For an intuition close in spirit, see Meibauer [2005].

from distinguishing lying from falsely implicating, when we assess types of cases *within* the domain of deception.)¹⁸

Our discussion showed that creating false beliefs through non-propositional means is manipulative without being deceptive. That analysis introduced the category of non-deceptive manipulation that causes false beliefs. We can now employ our new perspective to analyse the other ways of creating false beliefs. The next question to arise is this: when *true* propositions (in the sense specified above) create false beliefs, is this deception or instead manipulation?

As I presently discuss, such cases are instances of manipulation, not of deception. We ask: when we create false beliefs through communicating truths, by what mechanism do the false beliefs arise? Since these communicative acts convey *truths*, not falsehoods, their mechanism of action if they want to generate false beliefs must be by manipulating the way in which judgment works (obfuscating, deflecting, etc.), not by letting it function normally while feeding it wrong data, that is, not by convincing of falsehoods. It is therefore to be expected that all such cases will be manipulations. We shall now review salient schemes of misleading by communicating truths (providing an exhaustive list is, of course, impossible), and shall see how their analysis confirms our supposition.

First example. It is possible to intentionally create false beliefs by saying the truth, the whole truth, and nothing but the truth. This can be achieved, for instance, by choosing a rather expansive interpretation of what the *whole* truth consists of, and consequently by offering the victim such detailed information that it will predictably obfuscate understanding and result in false beliefs [Camerer, Loewenstein, and Weber 1989]. (Think, for example, of too detailed a disclosure of information in medical scenarios of informed consent.) It surely defies common sense to refer to the truth, the whole truth, and nothing but the truth as being deception. At the very least, we can affirm this: once we recognize the distinction between deception and non-deceptive manipulation that causes false beliefs, we cannot escape determining the precise border between them; and it is extremely intuitive to leave ‘the truth, the whole truth, and nothing but the truth’ outside of ‘deception’. Indeed, the mechanism here relies on the prediction that the information overload will impede the good functioning of judgment (a problem of its ‘form’, not its content); this defines manipulation.

Second example. False beliefs can be intentionally created while delivering precise truths, by using framing effects. Consider the following. By the entrance to the store, a little sign says that payment with credit card incurs a three percent surcharge. The shrewd vendor at the nearby store posts the *same* information by presenting it as a cash discount instead of a credit card surcharge, anticipating correctly that most people will develop the false belief that shopping at his store is cheaper [Thaler 1980]. The vendor provides accurate information while predicting that common cognitive biases will cause the receivers of the information to make wrong inferences. The mechanism of creating false beliefs here is not one of convincing anybody of a falsehood: if we claimed this, we would have to claim it about *both* shop owners equally (as they provided equivalent information only in different formulations); rather, they both provided blamelessly accurate and legitimate information. The second vendor exploited the public’s

¹⁸ Jennifer Saul [2012a] makes the important claim that linguistic classifications should be sensitive to the ethical aims in making such classifications—this is the intention here, too.

tendency for suboptimal use of judgment and insufficient reflection. Intentionally causing false beliefs by means of triggering the victims' lapses of judgment is the sign of manipulation. The use of the framing effect by the vendor is a paradigm of non-deceptive manipulation that causes false beliefs. To further enhance this insight, imagine that 95% of customers pay with credit cards; then the manipulative vendor's message is not only true, it is also the most natural form of presenting the truth. How could *that* be considered deception?

Third example. False beliefs are commonly created by coupling true messages with non-linguistic cues of various sorts. For instance, instead of simply telling you the fully accurate and unbiased information, I present it to you along with a melody, as in a jingle. This form of communication, notably used in advertising, is overwhelmingly pervasive in our culture. Despite the *ex hypothesi* accuracy of the information, this form of delivery can be used reliably to create false beliefs in the audience. The main mechanism is that of 'evaluative conditioning', in which (positive) attitudes toward an object are induced by classical conditioning (associating the object with positively valued unconditioned stimuli, such as upbeat music) [De Houwer et al. 2001]. The Pavlovian positive attitudes then generate corresponding unfounded beliefs about the object [Biegler and Vargas 2013]. Despite intentionally causing false beliefs, influences of this type are obviously not deceptions: the overt communication of fully accurate and unbiased information, whatever else it is, is not deception. It *is*, however, clearly manipulative—it is non-deceptive manipulation that causes false beliefs.

The examples above vindicate our analysis of the different mechanisms of deception and misleading manipulation. We conclude that creating false beliefs by means of true messages is most aptly classified as manipulation. If so, then communicating falsehoods is a necessary condition for deception.

4. Non-Deceptive Manipulation via Communicating Falsehoods

The next question, naturally, is that of whether communicating what is false is also a sufficient condition for deception: can there be intentional causation of false beliefs through communicating falsehoods that would not qualify as deception? If there cannot, the distinction between deception and manipulation would be neat indeed: creating false beliefs by means of communicating falsehoods would define deception; all other cases of creating false beliefs (that is, by non-propositional communication or by communicating truths) would be manipulations. I will now argue against that neat classification.

We started with two rival basic intuitions about deception—the intuition that it is to convince another of falsehoods versus the intuition that it is to cause another to have false beliefs. Our discussion has shown that understanding deception as the intentional causation of false beliefs fails miserably at sifting out non-deceptive manipulations that cause false beliefs. Moreover, we have learned that the *mechanism* of causing false beliefs is of decisive importance in classifying the phenomena. Having explained that deceptions necessarily communicate falsehoods, we have come very close to the first intuition; to vindicate it, we still need to provide an account of what 'convincing' the other of a falsehood amounts to. I will now argue for a straightforward interpretation—namely, that the misleading information must be the direct cause of the false belief. ('Direct cause' here refers to that which the listener takes to justify her belief. In manipulations, this could refer to various influences *other than* the falsehood told.)

To appreciate this, consider the following case. A tells B a falsehood p . P sounds too implausible to B; he therefore does not believe it. At the same time, B estimates that in the given circumstances it would be unreasonably risky for A to tell p if p weren't true, and he does not believe A to be such a bold risk-taker. Since the latter considerations outweigh the former, B comes to believe p . This entire course of events was premeditated and intended by A.

In this case, A intentionally tells p , which is false, to B, and B indeed ends up believing p . And yet A did not meet the criterion of *convincing* B of a falsehood: B emphatically did *not* believe A's assertion that p (the falsehood failed to cause a false belief). A did not therefore deceive B. True, in a roundabout way B did end up believing p ; however, it has already been established that successful intentional creation of false beliefs is insufficient to constitute deception—the mechanism is what makes the difference. The mechanism here is one of asserting falsehoods known to be unbelievable—a kin of 'bald-faced lies'—which does not constitute deception, but which here functions as a bait for B to misjudge a side issue (A's boldness). In the manipulation cases that we have reviewed, the proximal cause of the false beliefs was the victim's deficient judgment, not the communication of falsehood; here, too, the mechanism that generated the false belief was the misemployment of the victim's judgment. Since the entire course of events was premeditated and intended by the agent, this is a classic case of manipulation by inducing self-thwarting reflection in the other. (That the induction was through a falsehood cannot convert this into deception, as the falsehood itself was non-deceptive.)

Intent to create false beliefs, action in accordance with that intent, and the eventual creation of false beliefs are together insufficient to establish deception; the question of how these elements are interconnected is decisive. Consideration of the mechanism of creating false beliefs entailed that deception requires the communication of a falsehood; the extension of that same consideration supports the conclusion that the falsehood must, in addition, be the direct cause of the false belief created: the lack of veracity must be the agent of deception, the misleading factor. Otherwise, the direct agent of error would be the victim's own infelicitous judgment, and this—as we have consistently seen, starting with *Pill*—is what characterizes manipulations.¹⁹

Let us summarize. The condition of engaging the victim's agency in deception has too often been ignored; this arguably reflected the proclivity of moral philosophy to dwell on intentions and consequences at the expense of mechanisms. Once invoked, the condition of engaging agency is undeniable (remember the brain stimulation case), yet it demands clearer explication. Realizing this, non-trivial conclusions followed: deception requires the communication of a falsehood (in the extended non-technical sense articulated above), and it must directly cause the false belief. That this condition has not been recognized may be partly due to traditional philosophical literature's excessive focus on lying, where the falsehood communicated is, indeed, normally and trivially the agent of misleading. In other cases of producing false beliefs, we have to check whether the mechanism generating false beliefs is indeed the lack of veracity of

¹⁹ To the extent that the Gettier challenge to knowledge requires that the truth of the matter features as the direct justification of the true belief, in defining deception we find an 'inverse Gettier condition' of sorts: the falsehood communicated must feature as the direct justification of the false belief. (Further elaboration on this diagnosis must await another occasion.)

the message, or if it operates by triggering the victim's suboptimal use of his judgment. The former are deceptions; the latter are non-deceptive manipulations.

To the extent that philosophical analyses of deception traditionally focused on lying, the mostly undifferentiated basket of 'deception' has served as repository for anything not matching the definition of lying. The introduction into the picture of the (arguably more general) category of manipulations forces finer distinctions regarding the scope of 'deception'. The vantage point of this more synoptic view narrows substantially the category of deceptions.²⁰

We conclude that, no matter what the precise details of the definition of deception, it must incorporate a clause specifying that the false beliefs created are the direct result of being convinced of the falsehood communicated by the deceiver.

5. Moral Distinctions

What is the moral relevance of distinguishing deception from non-deceptive manipulation that causes false beliefs? According to a widely accepted view, certain ways of intentionally creating false beliefs are morally worse than others. The most prominent, classical, form of this intuition is that lying is morally worse than all other deceptions. This view is still prominent, yet contested (the notable alternative being that lying and other deception types are morally equivalent).²¹ Whatever the truth is in that debate, I propose here that the clearest, most significant, moral line between distinct ways of creating false beliefs is to be drawn not within deceptions but between deceptions and manipulations. There is a basic difference between the requirement to not communicate what is false and the requirement to ensure that the other's judgment (in response to the communicated message) is sound and therefore will not mislead her. The first is a negative duty, the second a positive one; and positive duties are either less stringent than negative ones or not duties at all. Hence, the positive 'duty' not to manipulate is either morally less stringent than the duty not to deceive or isn't a duty at all; in any case, these two differ in moral valence.

We must tread cautiously here, however. The validity and importance of the distinction between negative and positive duties is conditioned on our theories of what we owe to each other, as well as context-specific presuppositions of agentic responsibility, which can elevate positive action to the same level of moral requirement as that of negative duties. It is therefore important, when speaking here of 'positive' and 'negative', to give some richer background that supports the moral validity of the way in which the distinction is drawn. This gains special relevance, given that the distinction at which we are aiming *has* arguably been widely abused in moral thinking about deception. A prominent construal of what might be called the 'Augustinian-Kantian' school on deception suggests that the moral prohibition against lying is alone a negative duty, while not engaging in non-lying deception is merely a positive duty to assist others not to form false beliefs based on misleading conversational implicatures, as if such false beliefs are essentially the victim's business and responsibility. This construal runs counter to almost anyone's intuitions about truthfulness; it ignores the fact that pragmatic rules are as integral to standard linguistic practice as what is asserted. The logical

²⁰ Understanding deception as necessarily communicating falsehoods diminishes the gap between deceptions and lies, but major differences remain: lies alone cannot be non-linguistic and cannot assert what is literally true.

²¹ See Williams [2002] and Saul [2012b], regarding the equivalence between lies and false implicatures.

possibility of that construal, however, demonstrates why it is important to tell a richer story to support one's distinction between positive and negative duties. I will now argue that distinguishing morally between deceptions and misleading manipulations captures the grain of truth in the intuition that went amiss in the (above interpretation of the) 'Augustinian-Kantian' view.

What, then, is the background story that supports and confirms drawing the line of 'positive' versus 'negative' between manipulating that causes false beliefs and deceiving, respectively? While deception is *pro tanto* wrong, manipulation is emphatically not. Pervasive norms of human communication make it the case that many mundane manipulations are quite trivially morally innocent (we smile to ease social interactions, keep thoughts to ourselves, pick the opportune moment to make a request, etc.). If Erving Goffman's [1959] classical theory of social interaction (the 'dramaturgical approach') has even *some* truth to it, then normative social interaction could be described as *constitutively* manipulative. It is virtually impossible to imagine normal human interaction without subtle elements that—being used to influence people around us not through engaging their finer judgment—are best characterized as manipulation. (Regarding impossibility, think even of *Pill*. You might think it not too difficult to avoid manipulation by providing full disclosure regarding the potential effect of the colour blue. But this misses a central point: what is true for the pill's colour is similarly true for its size, the choice of brand name, the design of the case, the wording in the leaflet, and endless other cues. Full disclosure on all such matters would be so unreasonably tedious and confusing as to be *comic*.) A person who did not practise manipulation in any form would be, on the one hand, an utterly helpless creature (indeed, toddlers already manipulate) and, on the other hand, someone whose unfiltered candour would make her socially intolerable. Manipulation is thus a basic constituent of human well-being. A blanket moral condemnation would, accordingly, be unreasonable. Furthermore, a central function of everyday manipulation is to disrupt our transparency to others; achieving this through manipulations that cause false beliefs is indeed the overwhelming norm. Renouncing this protective mechanism could have devastating effects on the value of privacy, and thus could constitute an injury to human dignity. Moral theory, unless revolutionary and utopian, ought to accept all of this as a background condition for realistic theorizing. It is no surprise, then, that, unlike deception, manipulation is not *pro tanto* wrong.

Marcia Baron expresses an exact intuition in writing that 'benign manipulation is more nearly continuous with perfectly appropriate behavior, even valuable behavior, than is benign deception ... in some cases [of manipulation] there is nothing to justify; the manipulation ... not only is not wrong but is not even regrettable' [2014: 115–17]. This very fact constitutes a profound asymmetry that justifies drawing a moral line between deception and manipulation. Considering their respective mechanisms of action, this line can be described most naturally as separating negative duties from potential positive ones. ('Potential', since refraining from non-deceptive manipulation that causes false beliefs is often not a duty at all, as in the examples reviewed above.)

The overlooked distinction between deception and non-deceptive manipulation that causes false beliefs has been supported by three converging considerations. (1) It fits with common linguistic sensibilities: marketing pills in whatever colour is not normally considered deception, designating certain misleading manipulations—such as telling the whole and nothing but the truth—as deceptions is absurd, etc. (2) It relies on an account of different mechanisms of action. (3) The distinction drawn resonates with

sensible ethical intuitions regarding the (*pro tanto*) disparate levels of moral disapprobation between deceptions and misleading manipulations.

6. Ironic Cases of Intentionally Creating False Beliefs

The above analysis of misleading manipulations does not exhaust the contraction of the domain of deception. Considerations of conversational *context* further reduce its scope. Let us examine the nature of this further restriction.

Suppose that one creates false beliefs in another, in a way that we would normally count as deception. Now suppose further that the only alternative to creating those false beliefs would be to create other false beliefs instead. How should we think of this predicament? The creation of which false beliefs, if any, would count as deception? Think, for instance, of a case (commonly depicted in action movies) where a person undergoes such an exceedingly improbable series of events that there is no chance that any reasonable individual would believe his telling things just as they happened. He must therefore tell some untrue details if he sincerely wants his listeners to form a true belief about the general picture of what had in fact happened. The person relating the events has two alternatives: in both, he creates false beliefs, yet often none would be deception. The first alternative is to relate things exactly as they transpired, knowing that this will result in a false belief about the sheer occurrence of the entire event. Even if the speaker so acts while intending to create that false belief, it may not be deception—reporting exactly what happened is normally the *opposite* of deception. Most such cases would be yet another variant of non-deceptive manipulation that causes false beliefs via telling truths.²² Intentional creation of false beliefs is, again, not necessarily deception. We are, however, mostly interested in the second alternative, where the speaker intentionally does tell falsehoods about certain details, which directly create corresponding false beliefs. Normally, this would count as straightforward deception. I contend that, as long as two special conditions obtain, it nonetheless should not so count: (a) the speaker knows that, if he were to tell the truth about the details, his listeners would end up with false beliefs about the global story; (b) his telling the false details is based on good judgment that his particular listeners are interested not in the details but rather in the global story. In other words, I argue that, when causing false beliefs is both inevitable and meant to cause those false beliefs that are least misleading, even *convincing others of falsehoods* is no deception. In such cases, ironically, the agent must convince others of falsehoods for the sake of veracity.

Let us examine another example, exhibiting a similar logic, only here the twist is due not to the (lack of) believability but instead to the (lack of) sensibility of the events as they transpired. This similarly necessitates their being ‘edited for veracity’. Consider the following true testimony about the Assistant Secretary in the British War Cabinet during WWII whose job was to produce the Cabinet minutes, which were later praised for getting ‘as near the truth ... as anything could do’ [Gittings 1978: 70]:

First, his function was to take down by shorthand or notes what the ministers, or those summoned to advise them, actually said. Like all human discussion, perhaps especially in time of stress, this was virtually without form, incoherent, illogical, illiterate, and frequently

²² These cases have a structure similar to ‘double bluff’; and, while some double bluffs *are* deceptions, other cases sharing that structure are merely manipulations that cause false beliefs (elaborating on the distinguishing conditions cannot be attempted here). That *some* such cases aren’t deceptions suffices to make my point.

inconclusive. It was therefore the function of this official...to rewrite the whole affair, and give it the form it ought to have taken, to substitute in order, logic, and expression what he judged the participants had meant to say, not what they actually did say: for example, to make sense of a well-known explosive British general, who was so impatient of mere words that he often said China when he meant Russia, or Italy when he meant Germany

There was no way for that secretary to convey the truth about what really took place in the War Cabinet, without editing what was literally said. He had to introduce some untruths for veracity.

In the ironic cases of telling untruths for veracity, contextual considerations must enter to determine whether deception is the case, based on judging the *ad hoc* focus of conversational cooperation. If the agent relates false details in a context where the details specifically matter, this is deceptive. If, alternatively, the conversational context is characterized by an effort to understand the grand picture—as is so often the case in informational exchanges—then changing some details, as the inevitable means for achieving this goal, is not deception. The point is that to the extent that the intentional causation of false beliefs is the necessary vehicle for *optimal clarity in communication*, and is practised for that reason (namely, to *minimize misleading*), it is utterly counter-intuitive to call this ‘deception’. If the speaker has further intentions of influencing others by creating those beliefs, this could be yet another variant of non-deceptive manipulation that causes false beliefs; if not, it is not even that. The effect of context on the definition of deception is mediated through the idea of the cooperative goals of the communicative dynamic: while causing false beliefs in breach of those cooperative goals is deceptive, causing (other) false beliefs as a necessary means for conformity with those cooperative goals is not. In such ironic cases, convincing another of falsehoods can be a tool for *bona fide* conversational cooperation.

The intention to cause false beliefs and the intention to mislead have always seemed interchangeable explications of the intent to deceive. Here we see that they can part ways. When they do, common sense has it that it is the latter only that really defines the intent to deceive, and therefore deception. Hence, deception cannot be simultaneously compatible with the intent to minimize misleading. Deception is the act of intentionally misleading through convincing others of falsehoods. Without the misleading component, such acts may be mere manipulations, or not even that.

Intentionally causing false beliefs in ironic contexts teaches us a deep sense in which deception is ethically constructed. Traditional definitions of deception tend to separate the question of the nature of deception as departure from truth from the moral considerations pertinent to it. Thus, Chisholm and Feehan [1977: 143] commence their classic paper by declaring, ‘We are here concerned with what it is to intend to deceive and with what it is to lie. We will attempt to set forth a detailed analysis of these concepts ... We are not here concerned with ethical or moral questions.’ Carson [2009: 153] similarly starts his discussion by declaring, ‘Conceptual questions about the nature of lying and deception are prior to questions about the moral status of lying and deception.’ Our analysis has revealed, in contrast, that ethical considerations are not secondary, but rather participate in constructing the (borders of the) concept ‘deception’ via the idea of conversation as a cooperative enterprise. In one sense, this idea is presupposed by many conceptions of deception (through the notion of warrant of truth and corresponding breach of trust), but there is a significant difference between that sense and the sense presented here. In those other cases, cooperation is a general condition, embedded in the very meaning of ‘assertion’ or ‘conversation’. In our case, on the other

hand, being faithful to the cooperative enterprise requires an *ad hoc extra-linguistic* assessment of the point of the particular intersubjective interaction: we are asking not *whether* communication causes false beliefs (in our ironic cases, that is inevitable), but instead how relevant the false beliefs are to the aim of our particular interaction. This requires a special effort of attention and sensitivity to the interests of others in the reciprocal interaction. A failure of attention and sensitivity to the concerns of others is a moral failure. Such failure would undermine the successful appraisal of the real cooperative effort required, and would risk producing deception instead of mere irony. The imperative to avoid moral failure by sensitivity to the conversational goals of our interlocutors imparts some ‘thickness’ to the concept of deception.

7. Conclusion

Since any position in front of others is potentially communicative, the demand for veracity is among the most omnipresent of all moral demands. The precise scope of this demand is unclear, however. Disagreements refer not only to potential justifications for deception, but also to its very meaning. In a similar manner, manipulation is pervasive in virtually every corner of human interaction, yet both its meaning and moral justifications are quite obscure. This paper started with the observation that we can manipulate others into holding false beliefs without deceiving them. It subsequently introduced and analysed the category of non-deceptive manipulation that causes false beliefs, which purported to shed light on the elusive and largely neglected boundary between deception and manipulation. The distinctions that it helps to draw can, hopefully, contribute to a better understanding of both the ethics of deception and the ethics of interpersonal communication and influence, more generally understood.²³

References

- Baron, Marcia 2014. The *Mens Rea* and Moral Status of Manipulation, in *Manipulation: Theory and Practice*, ed. Christian Coons and Michael Weber, New York: Oxford University Press: 98–120.
- Biegler, Paul and Patrick Vargas 2013. Ban the Sunset? Nonpropositional Content and Regulation of Pharmaceutical Advertising, *The American Journal of Bioethics* 13/5: 3–13.
- Bizer, G.Y., J.C. Barden, and R.E. Petty 2003. Attitudes, in *Encyclopedia of Cognitive Science*, ed. L. Nadel, London: Nature Publishing Group: 247–53.
- Buss, Sarah 2005. Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints, *Ethics* 115/2: 195–235.
- Camerer, Colin, George Loewenstein, and Martin Weber 1989. The Curse of Knowledge in Economic Settings: An Experimental Analysis, *Journal of Political Economy* 97/5: 1232–54.
- Carson, Thomas L. 2009. Lying, Deception, and Related Concepts, in *The Philosophy of Deception*, ed. Clancy W. Martin, New York: Oxford University Press: 153–87.
- Carson, Thomas L. 2010. *Lying and Deception: Theory and Practice*, Oxford: Oxford University Press.
- Chisholm, Roderick M. and Thomas Feehan 1977. The Intent to Deceive, *The Journal of Philosophy* 74/3: 143–59.

²³ The author would like to thank Dalia Drai, Stephen Kershner, and Yakir Levin for their helpful comments on earlier drafts. Two anonymous reviewers for *AJP* deserve tremendous gratitude for their dedicated help in improving this paper greatly. Versions of this paper were presented in 2016 at the annual conference of the Israeli Philosophical Association, at Hebrew University in Jerusalem, the Rocky Mountain Ethics Congress at the University of Colorado, and the University of Hawaii–Manoa. Thanks for helpful comments are due to participants in all of these events, especially George Tsai from the University of Hawaii, and David Heyd from the Hebrew University in Jerusalem.

- Coons, Christian, and Michael Weber, eds. 2014. *Manipulation: Theory and Practice*, New York: Oxford University Press.
- De Craen, A.J.M., P.J. Roos, A.L. De Vries, and J. Kleijnen 1996. Effect of Colour of Drugs: Systematic Review of Perceived Effect of Drugs and of Their Effectiveness, *British Medical Journal* 313/7072: 1624–6.
- De Houwer, J., S. Thomas, and F. Baeyens 2001. Association Learning of Likes and Dislikes: A Review of 25 years of Evaluative Conditioning, *Psychological Bulletin* 127/6: 853–69.
- Faden, Ruth R. and Tom L. Beauchamp 1986. *A History and Theory of Informed Consent*, New York: Oxford University Press.
- Frijda, Nico, Antony S.R. Manstead, and Sacha Bem eds. 2000. *Emotions and Beliefs: How Feelings Influence Thoughts*, Cambridge: Cambridge University Press.
- Gittings, Robert 1978. *The Nature of Biography*, London: Heinemann.
- Goffman, Erving 1959. *The Presentation of Self in Everyday Life*, Garden City, NY: Doubleday.
- Gorin, Moti 2014a. Towards a Theory of Interpersonal Manipulation, in *Manipulation: Theory and Practice*, ed. Christian Coons and Michael Weber, New York: Oxford University Press: 73–97.
- Gorin, Moti 2014b. Do Manipulators Always Threaten Rationality? *American Philosophical Quarterly* 51/1: 51–61.
- Herrmann, A., M. Zidasek, D.E. Sprott, and E.R. Spangenberg 2013. The Power of Simplicity: Processing Fluency and the Effects of Olfactory Cues on Retail Sales, *Journal of Retailing* 89/1: 30–43.
- Krosnick, J.A., L.J. Jussim, and A.R. Lynn 1992. Subliminal Conditioning of Attitudes, *Personality and Social Psychology Bulletin* 18/2: 152–62.
- Mahon, James E. 2007. A Definition of Deceiving, *International Journal of Applied Philosophy* 21/2: 181–94.
- Meibauer, Jörg 2005. Lying and Falsely Implicating, *Journal of Pragmatics* 37/9: 1373–99.
- Mitchell, C.J., J. De Houwer, and P.F. Lovibond 2009. The Propositional Nature of Human Associative Learning, *Behavioral and Brain Sciences* 32/2: 183–98.
- Noggle, Robert 1996. Manipulative Actions: A Conceptual and Moral Analysis, *American Philosophical Quarterly* 33/1: 43–55.
- Saul, Jennifer M. 2012a. *Lying, Misleading, and What Is Said: An Exploration in Philosophy of Language and in Ethics*, New York: Oxford University Press.
- Saul, Jennifer M. 2012b. Just Go Ahead and Lie, *Analysis* 72/1: 3–9.
- Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.
- Thaler, Richard 1980. Toward a Positive Theory of Consumer Choice, *Journal of Economic Behavior & Organization* 1/1: 39–60.
- Todd, Patrick 2013. Manipulation, in *The International Encyclopedia of Ethics*, ed. Hugh LaFollette, Oxford: Wiley-Blackwell: 3139–45.
- Tsai, George 2014. Rational Persuasion as Paternalism, *Philosophy & Public Affairs* 42/1: 78–112.
- Williams, Bernard 2002. *Truth and Truthfulness: An Essay in Genealogy*, Princeton: Princeton University Press.