

**Self-deception and self-knowledge: Jane Austen's *Emma* as an  
Example of Kant's Notion of Self-Deception**

***Autodecepción y autoconocimiento: "Emma" de Jane Austen  
como ejemplo de la noción kantiana de autodecepción***

JEANINE M. GRENBORG\*

St. Olaf College Northfield, MN USA

**Abstract**

In this paper, I address the theme of harmony by investigating that harmony of person necessary for obtaining wisdom. Central to achievement of that harmony is the removal of the unstable, unharmonious presence of self-deception within one's moral character.

**Keywords**

Kant; Harmony; Self-deception; Self-knowledge; Wisdom

**Resumen**

Este artículo plantea la cuestión de la armonía de la mano de la investigación de la armonía que la persona precisa para obtener sabiduría. Sostengo que la supresión de la presencia inestable y no armónica del autoengaño en el marco del carácter moral de cada cual es central para alcanzar tal armonía.

**Palabras clave**

Kant; armonía; autoengaño; autoconocimiento; sabiduría

---

\* Professor of Philosophy of Saint Olaf College. E-mail for contact: [grenberg@stolaf.edu](mailto:grenberg@stolaf.edu).

I have, in the past, argued versus Onora O’Neill that knowledge of oneself (and not just right action) is a central moral concern for Kant, suggesting that a balanced amount of introspection was not only permissible but obligatory even to hope to fulfill one’s duties to others.<sup>1</sup> In this paper, I further my thoughts on this topic by considering what one might take to be the direct opposite of self-knowledge: self-deception. To deceive oneself is not only to fail in self-knowledge, but also to seek actively to avoid it. I have come to believe that Kant not only believes self-deception is possible but also that it is the most *characteristic* obstacle to successful expression of finite practical reasoning. If this is true, the obvious conclusion that follows is that the *removal* of self-deception, that is, self-knowledge is, in Kant’s words, “the **First Command** of All Duties to Oneself.” (6:441/191)<sup>2</sup> Indeed, if we accept Kant’s related point that duties to *self* are previous to all duties to *others*, then we arrive at the interesting conclusion that self-knowledge is the first of all duties *simpliciter*. I accept these strong claims, and seek here to defend them.

### 1. *Self-Deception in the Metaphysics of Morals*

Kant investigates self-deception in his discussion of lying in the *Metaphysics of Morals*. There, he seems perplexed by the phenomenon. On the one hand, “[i]t is easy to show that man is actually guilty of many inner lies.” (6:430/183) Kant accepts as obvious that lying to oneself is prevalent in humanity, and that such a tendency is easily observable. He quickly raises a problem, though: “it seems more difficult to explain *how* [inner lies] are possible; for a lie requires a second person whom one intends to deceive, whereas to deceive oneself on purpose seems to contain a contradiction.” (6:430/183). The implicit “contradiction” is that I need at once to know my intention to lie (as the “liar”) but also *not* to know my intention to lie (as the “lied-to”). So the perplexing thing about self-deception is that inner lies occur, but the very notion of an inner lie involves a contradiction which seems to make it impossible for them to occur.

Although Kant lays out this difficult question, he does nothing—at least in the *Metaphysics of Morals*—to resolve it. Instead, he dwells on the unquestionable phenomenon of self-deception in several examples. I’m not certain whether all his examples really are examples that fit the model he gives, viz., that each of these persons both knows and does not know his intention to lie. For example, he suggests that one who “professes belief in a future judge of the world, although he really finds no such belief

---

<sup>1</sup> See: Grenberg, Jeanine. *Kant and the Ethics of Humility: A Story of Dependence, Corruption and Virtue* (Cambridge: Cambridge University Press, 2005).

<sup>2</sup> All reference to Kant will first reference the Akademie edition page numbers, followed by the page numbers of the following Cambridge translation of Kant’s works:

*Critique of Pure Reason*, Paul Guyer and Allen Wood, trs. (Cambridge: Cambridge University Press, 1999).  
*Groundwork of the Metaphysics of Morals*, Mary Gregor, trs. (Cambridge: Cambridge University Press, 1998).

*The Metaphysics of Morals*. Mary Gregor, trs. (Cambridge: Cambridge University Press, 1996).

within himself” is someone who is deceiving himself. But this person seems more of a straight-forward prudential reasoner than a self-deceiver. He “persuade[s] himself that it could do no harm [to profess a belief in God]” and that it “might even be *useful* to profess in his thoughts...a belief in such a judge, in order to win his favor in case he should exist.” (6:430/183, emphasis added) This describes someone who knows very well that he does not believe in God, but who is going to pretend to believe in God just in case God does exist. This is just prudential reasoning (and outright deception of *others*), not self-deception. Mind you, it is *bad* prudential reasoning, since if God really were to exist, He would be the first one to see through this little charade. Nonetheless, this is not a good example.

Kant’s second example is better: “Someone also lies if, having no doubt about the existence of this future judge [i.e., God], he still flatters *himself* that he inwardly *reveres* his law, though the only incentive he *feels* is fear of punishment.” (6:430/183, emphases added) Here, the lie to oneself is that one feels reverence toward God when in fact one only feels ‘fear of punishment’, not ‘reverence.’

Let’s speculate now on the *motives* for this self-lie. Why would one want to believe that one is a reverent and religious person when in fact one is only a fearful, sycophantic panderer to a powerful being? Put that way, the motive for self-deception is screamingly obvious: one simply does not want to believe that one is as horrible a person as one has turned out to be. In Kant’s words, this person wants to avoid the “descent into the hell of self-cognition.” (6:441/191) (He also calls this “the abyss” of self-knowledge – 6:441/191) None of us likes admitting parts of ourselves that are less-than-good. So, instead of admitting we are less-than-good and then working at making ourselves better, we hide from ourselves that we are less-than-good and then present ourselves as even better than we are! This tendency toward false, arrogant self-presentation is one of the most frequent fruits of self-deception. Self-deception becomes a tool whereby we construct an image of ourselves more pleasing to ourselves. We then go out into the world with that constructed image, building an entire reality around us that supports the original self-lie. I may really not have reverence toward God, but I ostentatiously present myself to everyone as though I do; I even come to believe that I am *truly* reverent toward God. If I’m lucky, people start saying of me that I am an exceedingly reverent person. More likely, people will think I am a very hypocritical person. But the hypocrisy at the basis of my character is exactly what I am preventing myself from knowing about myself.

This tendency to “live the lie” is one reason Kant thinks self-deception is such a bad thing. He states:

«[S]uch insincerity in his declarations, which a human being perpetuates upon himself...deserves the strongest censure, since it is from such a rotten spot (falsity, which

seems to be rooted in human nature itself) that the ill of untruthfulness spreads into his relations with other human beings as well» (6:430-431/182)

Once we lie to ourselves we cannot help but to lie to others (in the sense of falsely presenting ourselves to them), and what started as one wrong perpetuates a whole series of wrongs, indeed, for some, a whole lifetime of wrongs towards oneself and others.

An excellent example of Kant's point about how self-deception perpetuates wrongs can be found in Jane Austen's novel, *Emma*.<sup>3</sup> Emma is in a good situation at her family estate, Hartfield. Her mother's early death assured that she became lady of the manor early on. She is "handsome, clever and rich," and lives comfortably with her "affectionate, indulgent" father. (Austen, 5) Mr. Knightley is a frequent visitor and close friend of the family, so Emma has everything she needs.

All this inspires Emma's main motive for self-deception: because she is frightened to lose her happy situation at Hartfield, she constructs a belief that she never wants to marry. Marrying would, after all, require her to leave Hartfield, her father and visits from Mr. Knightley. She thus deceives herself into believing both that she is not in love with Mr. Knightley, and that she does not want to marry. Emma also falsely believes, and takes great pride in the belief, that she is an accomplished match-maker, a false belief rooted in her unwillingness to admit to herself that she is a rather lazy person, and isn't good at the sort of things—like painting or music-making—which actually *do* take time, hard work and discipline. Emma thus constructs a world around her which supports all these false beliefs.

Emma's self-deception on these points leads her utterly to misunderstand expressions of affection toward her from men. One of my favorite passages for appreciating this point is when Emma, in matchmaking mode, is trying to convince her friend Harriet that a riddle written by Mr. Elton is in fact about his love for Harriet:

«She cast her eye over it, pondered, caught the meaning, read it through again to be quite certain...and then passing it to Harriet, sat happily smiling... 'May its approval beam in that soft eye!' Harriet exactly. Soft, is the very word for her eye—of all epithets, the justest that could be given. 'Thy ready wit the word will soon supply.' Humph— Harriet's ready wit! All the better. A man must be very much in love indeed, to describe her so» (Austen, 58).

The reader is clearly meant to realize this poem about a woman with a "soft eye" and "ready wit" is Emma. But Emma's self-deception about her own interest in love leads her to misinterpret it as being about Harriet, one more with dull eyes than soft ones and a dull wit (rather than a 'ready' one) to match them.

---

<sup>3</sup> All references to *Emma* come from: Austen, Jane. *Emma* (Oxford: Oxford University Press, 2008).

Emma's self-deceptions have exactly the results Kant suggests: they are a 'falsity...that...spreads into [one's] relations with other human beings,' (6:430-431/182) which guarantees that wrongs are brought upon others. Not only does Emma inadvertently present a mistruth to Harriet about Mr. Elton's affections. Beyond that, Emma's encouragements raise Harriet's hopes. So, when the truth of his affections is revealed, Harriet's pain is intense, and Emma has been the cause of that pain.

So, if self-deception is 'rooted in human nature itself' and has this huge moral consequence of grounding a false life, injurious to ourselves and others, we can see why getting to *know* oneself—and especially uncovering deceptions at the basis of one's self-presentation—is a central moral task. In a section titled "the **First Command** of All Duties to Oneself," (6:441/191) Kant affirms:

«Moral cognition of oneself, which seeks to penetrate into the depths (the abyss) of one's heart which are quite difficult to fathom, is the beginning of all human wisdom. For in the case of a human being, the ultimate wisdom, which consists in the harmony of a human being's will with its final end, requires him first to remove the obstacle within (an evil will actually present in him) and then to develop the original predisposition to a good will within him, which can never be lost. (Only the descent into the hell of self-cognition can pave the way to godliness.)» (6:441/191)

If we were not the sort of persons to deceive ourselves, then self-knowledge might not be so important. Getting to know ourselves would be like getting to know anything we have not yet understood, like calculus or flute-playing. But the duty to self-knowledge is a duty to know 'evil' things about you which you have actively sought to avoid knowing, things which require moral *attention*, not avoidance. Acquisition of self-knowledge puts us on the road to 'wisdom' (and not just 'knowledge') because it removes that internal obstacle which prevents us from developing the goal of a wise, moral person: a will in 'harmony' with a good will (instead of one actively opposed to that end). Self-deception is a tool for false character building. But self-knowledge is an act of wisdom because, through it, we uncover that false character. It brings out into the open our self-incurred obstacles to morality, our efforts to *hide* from ourselves what really needs to be *worked* on to become a moral person.

Scrutinizing our motives and character is not, however, an end in itself. The whole point of admitting who one has *been* is to become who one was *meant* to be. All those lies to oneself turned back on the world as arrogant self-assertions must be undermined so that new, self-aware motives can provide the basis for a new character in 'harmony' with a good will. The 'wisdom' here is of one who *knows* that the unexamined self is the first and worst obstacle to becoming moral. Indeed, Kant would, with Socrates, not only agree that

the unexamined life is not worth living, but also add that the *unexamined* life is *guaranteed* to be an *immoral* life.

## 2. Explaining How We Deceive Ourselves

### 2.1 Step One: Groundwork I

Yet, even as we draw this picture of self-deception, we still don't know how it is possible. There remains this apparent "contradiction" that the same person is both liar (and thus aware of her intention to deceive) and lied-to (and thus unaware of her intention to deceive).

Although we get no resolution of this contradiction in the *Metaphysics of Morals*, one brief hint in his discussion there points us forward. After raising the problem of contradiction, Kant distinguishes between "[m]an as a moral being" and man "as a natural being", suggesting that, like all persons, "[m]an as a moral being...cannot use *himself* as a natural being...as a mere means (a speaking machine)...but is bound to the condition of using himself as a natural being in agreement with the declaration...of his moral being and is under obligation to *himself* to truthfulness." (6:430/183) The implied thought here is that self-deception somehow involves one person having two selves: the "moral being" and the "natural being," with the former admonished not to treat the latter as a mere means toward the end of his lying.

Kant gives no further hints of how to make sense of the interaction between these two beings within one person. But, if we turn to *Groundwork I*, we find a story of attentiveness, and failure thereof, which constructs a clearer picture of the interaction of two internal selves. Along the way, we'll find that self-deception is deeply embedded in the nature of human practical reasoning itself and is thus *the* defining, characteristic challenge to becoming a moral person.

At the very end of *Groundwork I*, Kant provides an account not only of *why* we deceive ourselves about moral obligations, but also hints of *how* we do so:

«The human being feels within himself a powerful counterweight to all the commands of duty, which reason represents to him as so deserving of the highest respect – the counterweight of his needs and inclinations, the entire satisfaction of which he sums up under the name happiness. Now reason issues its precepts unremittingly, without thereby promising anything to the inclinations, and so, as it were, with disregard and contempt for those claims, which are so impetuous and besides so apparently equitable (and refuse to be neutralized by any command). But from this there arises a *natural dialectic*, that is a propensity to rationalize against those strict laws of duty and to cast doubt upon their validity, or at least upon their purity and strictness, and where possible, to make them better

sued to our wishes and inclinations, that is, to corrupt them at their basis and to destroy all their dignity» (4:405/17-18).

This person is in the process of deceiving himself about the authority of moral demands. The truth of the matter is that moral demands are categorical. But, here, he begins to turn them into something else—laws that are perceived *not* to hold with categorical ‘purity’ or ‘strictness’—via self-deception.

First, let’s think about *why* the person engaged in this ‘natural dialectic’ wants to deceive himself. He wants to place happiness *above* morality when the two conflict; so deceiving himself about the strictness of the moral law becomes an attractive option. This prioritizing of happiness over morality is just the most general characterization we can give of the motives of *any* case of self-deception. The man who wants to believe himself reverent toward God even though he’s not wants to avoid the injury to his happiness that admitting his baser motives would involve. All Emma’s self-deceptions are pointed toward retaining her happiness at Hartfield. Now, we see the same desire for happiness encouraging this man to deceive himself about the strictness of moral demands.

*How* though does this man deceive himself? The crucial moment for understanding the mechanics of self-deception is when this man ‘cast[s] doubt upon [moral laws]’ validity, or at least upon their purity and strictness, and where possible...make[s] them better suited to [his] wishes and inclinations.’ To make sense of this move, let’s go back just a page earlier in *Groundwork I* to remind ourselves that Kant first *affirms* the capacity of “common human understanding” to get things *right* here, that is, to see *clearly* the ‘strict’, categorical nature of moral laws:

«[C]ommon human reason...knows very well how to distinguish in every case that comes up what is good and what is evil, what is in conformity with duty or contrary to duty, *if*, without in the least teaching it anything new, we only, as did Socrates, *make it attentive to its own principle*; and that there is accordingly, no need of science and philosophy to know what one has to do in order to be honest and good, and even wise and virtuous0 (4:404/16, emphases added).

*As long as* one pays attention to the presence of moral demands within one’s moral consciousness, their strictness is very clear. In the *Critique of Practical Reason*, Kant confirms this import of attentiveness. There he notes: “[w]e can become aware of pure practical laws...by *attending* to the necessity with which reason prescribes them to us and to the setting aside of all empirical conditions.” (5:30/27, emphasis added) Here, not only do we appreciate *that* the person of common reason attends to the moral law; we also learn more about *how* he does it: by ‘setting aside...all empirical conditions,’ that is, by not letting desire-based or happiness-focused interests interfere with one’s perception of the moral law.

Now, back in the *Groundwork*, the person of “common human reason” (Kant now simply calls him “the human being” – 4:405/17) starts to “attend” rather differently to things. If he looks clearly at the moral law, setting aside all concerns about happiness, he knows it for exactly what it is. But our ‘human being’ is now engaged in a conflict between happiness and morality, and he wishes happiness could win. Well, then, let’s not look at morality so clearly. Let’s “attend” to something else so that the law looks differently to us. That “something else”, of course, is our own happiness and whatever oddly colored lens it provides through which to see the moral law. “Of *course* one shouldn’t lie, but maybe just this one time for the sake of paying my debt off (which will make me happy). I’ll even make other people happy because I’ll stop pestering them with further requests for loans. Yeah, that’s it: I’m really *helping* other people if I lie now. It really is the right thing to do.” If I attend to my happiness as primary instead of attending to the strict demand of the law, the law starts looking, well, less strict. I may still (in some sense to be considered later) “know” that the moral law holds strictly, but I no longer *believe* that. I no longer *attend* to that knowledge, so it recedes quietly into the back of my mind and a new belief emerges: the moral law *isn’t* totally strict; it admits of exceptions in the name of my happiness. Kant had told us the attentive human being needed only to ‘set aside...all empirical conditions’ to see the moral law clearly. But this person does exactly the opposite: he *welcomes* ‘empirical conditions’ into his perception of the law, and thus deceives himself about it.

To summarize: our person reflecting on the moral law does not stop knowing that it holds categorically; but he develops techniques to stop *looking* attentively at that knowledge. He looks elsewhere—to his own hopes for happiness—to gain a new, and corrupting, perspective on the moral law. Doing so allows him, eventually, perhaps over a long period of time, to begin *believing* the precise *opposite* of what he already *knows* to be true (viz., the categorical authority of the moral law). So, this person *believes* both a and ~a: the moral law holds categorically, and the moral law does not hold categorically. But he develops a disposition, or way of being, in which he *attends* only to ‘~a’ (the moral law does not hold categorically). And he attends to that so regularly over time that he begins to *lose* his *conscious* memory of ‘a’ (the categorical authority of the moral law).

## 2.2 Step Two: Analyzing Two Selves via Emma and Groundwork I

Can we, then, appeal to these shifting objects of attentiveness to explain how the same person is both the liar (who knows the intention to deceive) and the lied-to (who doesn’t know that intention)? Well, in fact, we’ve uncovered in this belief of ‘a’ and ‘~a’ story a slightly different “contradiction” within self-deception than the original one Kant had suggested. Nonetheless, investigating further how the construction and management of this contradiction in knowledge claims occurs *will* allow us to affirm, to a certain extent, the other “two selves” story Kant suggested in the *Metaphysics of Morals*. We need, then,

to investigate further this person contradictions in beliefs and with that, construct a clearer account of the two selves involved in self-deception.

First, the knowledge we attribute to this person of ‘a’ (the moral law holds categorically) is only implicit. He doesn’t *claim* the moral law holds categorically; indeed, he might become a defender of the idea that one *must adjust* the strictness of moral laws in light of happiness. But he can hold this claim implicitly only within a carefully constructed world maintained by a careful segmenting of himself into two. We had been saying these two selves are the liar (who knows the intention to lie) and the lied-to (who does not). But the self does not segment *exactly* along these lines. Rather, there is, first, the self able to look at truths directly and know the strictness of the moral law; let’s call him the Honest Self. Then there is the self who wants to satisfy the demands of happiness over those of morality. This is the Devious Self. The interesting thing, though, is that the Honest Self doesn’t become “the lied to” and the Devious Self doesn’t know himself constantly as the liar. Something different occurs. Let’s investigate each of these points in turn.

First, the Honest Self is not told a lie by the Devious Self. Rather, the Honest Self just gets ignored, sort of lost in this person’s overall consciousness: he has his knowledge, but he has a curtain placed over him by the Devious Self, who won’t pay attention to him. He is no longer involved in the choice process of the self, at least not directly.

On the flip side, the Devious Self, who *puts* the curtain over the Honest Self, cannot be said to be continually aware of the intention to deceive. That’s why I call him the Devious Self instead of the Lying Self. If we think of self-deception as a *process* through which one eventually comes to believe the opposite of what one knows, then the Devious Self, instead of maintaining a continual awareness of his intention to deceive, looks more like this: initially, he has an intention to lie (i.e., to claim ‘~a’ instead of ‘a’). But stated so baldly, this is too much: make moral laws bend to the demands of *my* happiness?? That would be wrong!! So, the Devious Self starts to build a world within which that obviously wrong thing no longer *looks* so obviously wrong. Now, ‘~a’ doesn’t look so bad. It doesn’t even look like a “lie” anymore. So the Devious Self *claims* ‘~a’, but he doesn’t claim it (even to himself) *as* a “lie”; it is just his way of seeing the world. So, he does *not* maintain a conscious *intention* throughout this process to lie to the Honest Self, or to anyone else. It might have begun like that, but by the time the deception is complete, it no longer looks like that.

And yet the Devious Self *does* “attend”—and attend very *carefully*—to the beliefs of the Honest Self, i.e., to ‘a’ (‘the moral law holds categorically’). How? Only in the back-handed sense of assuring vigorously that the truth of the Honest Self *not* be revealed,

especially in those circumstances in which such outing appears immanent. Let's look again at *Emma* to appreciate this point.

A dangerous possibility of outing Emma's false beliefs occurs when Mr. Knightly tries to call out the Honest Emma, insisting that she is in fact not a 'successful' matchmaker at all:

«I do not understand what you mean by "success;"... 'Success supposes endeavor. But if, which I rather imagine, your making the match, as you call it, means only your planning it, your saying to yourself one idle day, "I think it would be a very good thing for Miss Taylor if Mr. Weston were to marry her," and saying it again to yourself every now and then afterwards, --why do you talk of success? Where is your merit? --what are you proud of? --you made a lucky guess; and *that* is all that can be said» (Austen, 11).

Here is an excellent opportunity for the Honest Emma to come out: "oh, alright, Mr. Knightly, my matchmaking is all just a front for my failure to admit that I am lazy." Well, she doesn't say *that* in response to Mr. Knightley. Instead, when Mr. Knightley concludes "you made a lucky guess; and that is all that can be said," Emma responds:

«And have you never known the pleasure and triumph of a lucky guess?—I pity you.—I thought you cleverer—for depend upon it, a lucky guess is never merely luck. There is always some talent in it. And as to my poor word 'success,' which you quarrel with, I do not know that I am so entirely without any claim to it. You have drawn two pretty Opictures—but I think there may be a third—a something between the do-nothing and the do-all. If I had not promoted Mr. Weston's visits here, and given many little encouragements, and smoothed many little matters, it might not have come to anything after all» (Austen, 11).

This is Emma's Devious Self attending to her Honest Self's knowledge, and coming to protect it from being revealed. To admit to Mr. Knightley that she is lazy and unaccomplished would be too much. But her Devious Self is well-trained in such matters, and gets right to work. First, she speaks in an arrogant, almost insulting, tone ('I pity you' and 'I thought you were cleverer'), hinting at some defensiveness underneath her words. She tries, furthermore, to claim there was hard *work* behind her match-making success, thereby affirming the world the Devious Self has constructed.

This arrogant assertion of the world of the Devious Self confirms, then, one sense of the self-deceived person's "knowledge" of the denied fact: Emma knows that she is lazy and unaccomplished in the sense that she assiduously prevents this fact from coming to light, either to herself or others. When circumstances threaten to 'out' that fact, her Devious Self is immediately aware of the threat and fiercely protects that fact, keeping it in its safe hiding place deep within her curtained Honest Self. The denial of the guarded fact finds its strength in the intense desire to keep hidden what one wants to hide. As such,

“knowledge” of the denied thing must be admitted as a condition for the very need for, and the resulting strength of, the denial.

This “knowledge” of the underlying truth when it presents itself thusly is, however, a complex state psychologically. Long ago, Emma just knew that she was lazy and unaccomplished. But now, at the same time, she holds all the following:

- 1) She unconsciously, but accessibly, *knows* that she is lazy and unaccomplished.
- 2) She *believes* she is *not* lazy and unaccomplished.
- 3) She *fears* she *is* lazy and unaccomplished.
- 4) She does not *want* to believe she is lazy and unaccomplished.
- 5) She fiercely *denies* any claims that she is lazy and unaccomplished.

Emma holds all these states, but *attends* differently to each, depending upon circumstances. When Mr. Knightly suggests she is not accomplished in matchmaking, she most likely starts, internally, to attend to #3 (her *fear* that she *is* lazy and unaccomplished), leading to conscious assertion of #5 (her denial that she is lazy). In this state, her knowledge of #1 (that she is lazy and unaccomplished) remains unconscious.

Ironically, though, the knowledge of the Honest Self, despite being unconscious, is at the very *basis* of the plotting of the Devious Self, acting as that part of the self which must never be explicitly revealed. That truth is “known” in the sense of being that which must always be denied. The man of *Groundwork I* thus “knows” in his Honest Self that the moral law holds strictly, but it is too painful a fact to admit to himself or others. His Devious Self’s way of “knowing” this fact is to fiercely protect it, keeping it in its safe hiding place deep within his curtained Honest Self. The contradiction here is not that he both knows and does not know the intention to lie. The contradiction is that, in different ways, he believes both *a* (the moral law holds strictly) and  $\sim a$  (the moral law does not hold strictly); it is not the contradiction that he both knows and does not know the intention to lie.

But, although we thus reject the contradiction within the self-deceived person Kant had originally suggested, we *affirm* Kant’s other *Metaphysics of Morals* distinction between a moral self and a natural self. More precisely, the Devious Self fails to attend to the moral admonition the “moral self” was given. Kant had said the moral being must not use the natural being as a mere means. I’m not sure my Devious Self and Honest Self are perfectly parallel to his moral being and natural being. Nonetheless, the Devious Self is *using* the knowledge of the Honest Self as a mere means to his end of securing his happiness. Whether he *intends* to lie or not, the Devious Self is *using* another part of himself as a mere means instead of accepting that part of himself—the part that knows the truth of things—on its own terms, or as possessing its own “end” of truth-telling.

It is, of course, at least theoretically possible that, instead of continuing to conceal the Honest Self, the Devious Self will just give up and admit what had been hidden. But, such direct attentiveness to the fact to be denied would make the entire structure of self-deception collapse into the simple state of direct knowledge of the previously denied fact. Such reactions are thus rare, since they indicate the barest beginnings of that road to self-knowledge of which Kant spoke. They are, however, possible. The most likely scenario within which honest admission of the previously denied truth would emerge is one in which the original intentions of the Devious Self are thwarted. Let's turn again to *Emma* to appreciate the point.

First, some background: the unfortunate target of Emma's matchmaking attention is Harriet Smith, a woman younger than Emma, a little inept, and with lower social standing. Emma engages in two spectacular failures of matchmaking when she tries to connect Harriet with two different men who in fact are attracted to Emma, not Harriet. But the beautiful, bitter irony of Austen's narrative is that the person Harriet eventually does fall in love with is...Mr. Knightley, *Emma's* Mr. Knightley!!

When faced with the real possibility that someone else could be attracted to *him* and that (gasp!) *he* might be attracted to someone other than herself, Emma is finally forced to admit that she *does* love Mr. Knightley. Emma is *not* welcoming that path to wisdom which is self-knowledge; she is, rather, being untimely forced to admit things about herself as she watches the hope of her undisturbed life at Hartfield being challenged by little Harriet, whom she thought to be her toy, her protégé, and who now is revealing herself more as a monster in the Frankenstein sense, whose capacities exceed those of her creator's!! This is a perfect example of how changes in circumstances thwart the plotting and planning of the Devious Self; for, surely, Mr. Knightley marrying someone else would destroy Harriet's happy world at Hartfield, and it is the protection of that world which inspired Emma's self-deception in the first place.

Here, then, is the moment when she gives up what she had previously vociferously denied, viz., that she loves Mr. Knightley:

«To understand, thoroughly understand her own heart, was the first endeavor...How long had Mr. Knightley been so dear to her, as every feeling declared him now to be? When had his influence, such influence begun?—When had he succeeded to that place in her affection...?...—She saw that there never had been a time when she did not consider Mr. Knightley as infinitely the superior, or when his regard for her had not been infinitely the most dear. She saw, that in persuading herself, in fancying, in acting to the contrary, *she had been entirely under a delusion, totally ignorant of her own heart* (Austen, 324, emphasis added)».

We see here the other sense in which we can say the self-deceived person “knows” the denied truth: given the right circumstances—and especially given circumstances in which the efforts of the Devious Self are foiled—the truth that has been denied by the self-deceiving person comes naturally to light. Of course she has always “known” that she loved Mr. Knightley; she just couldn’t admit it to herself. Her knowledge of the previously hidden fact is now clear: she simply has conscious knowledge of it.

I suspect that if our *Groundwork* I man rejecting the strictness of moral laws for the sake of his own happiness were, ironically, presented with a similar sort of situation—one within which holding that claim (that moral laws are *not* strict) were to *undermine* his happiness, then he’d abandon his constructed world and admit that moral laws *do* hold strictly. Suppose, for example, that someone else uses a non-strict reading of moral laws to cheat him. Our man would vociferously object that moral laws *do* hold strictly! I do not mean to suggest this man will become moral (certainly not for the right reasons), but he does see his original project of self-deception failing; there is a collapse of the structure of the world he had created for himself because he now has to admit, *for* the sake of his own happiness, that moral laws *do* hold strictly!

So, we have two examples of the senses in which we can say that the self-deceived person “knows” the truth of the thing about which he’s been deceiving himself, either indirectly through vigorous denial of the truth, or directly, leading to the collapse of self-deception. And what we have accomplished in analysis of Emma’s self-deception, instead of a resolution of the original contradiction at the basis of self-deception, is an *affirmation* of the *existence* of a slightly *different* sort of contradiction hiding underneath an unstable state of self-deception. The self-deceiving person is not both a liar who knows the intention to lie and a lied-to who does not know that intention. Rather, the same person believes ‘a’ & ‘~a’, but *manages* those contradictory beliefs so as to efface the power of the contradiction. This management is accomplished by segmenting the self into the Devious Self and the Honest Self, where the Honest Self becomes used as a mere means by the Devious Self.

In this management of selves, we also see the management of various knowledge claims and beliefs. Some beliefs (as, e.g., the belief that Emma is *not* lazy and unaccomplished) are conscious; other beliefs (as, e.g., the underlying belief that Emma *is* lazy and unaccomplished) are normally unconscious, but accessible, depending upon circumstances. Usually, though, this underlying knowledge/belief is relegated to the realm of the “forbidden,” that of which one does not speak and does not even consciously know. As long as the Honest Self enters into the Devious Self’s projects only in the way the Devious Self wants, management of one’s contradictions succeeds: the same person can indeed successfully believe both a & ~a! There *is* a contradiction at the basis of self-deception.

But this need for complex management of one's contradictions is also what makes self-deception such an unstable state. Once one has to deny the '¬a' claim of the Devious Self (as Emma is forced to do when she is forced to admit her love for Mr. Knightley), the whole managed structure of self-deception comes crashing down: *without* a contradiction carefully held and managed, there *is* no self-deception.

We thus do not so much *resolve* the contradiction inherent to self-deception, as instead *affirm* it. The difficult work of the self-deceived person is to manage her holding of contradictions in belief in a way that makes them less obviously contradictions. It is too simple merely to say that one part of the person is the liar who knows the deception, and the other part is the lied-to who is deceived. The liar turns out to know her intention to lie only before she turns that intention into something else; and the lied-to is never really the lied-to but only the ignored, yet protected and used.

### 3. Conclusion

There is, however, a final point to make about self-deception, and ultimately, self-knowledge, for Kant. When discussing self-deception in *Groundwork I*, Kant spoke of it as a "natural dialectic," a phrase that should prick the attuned ears of a Kant scholar. Recall that a "natural dialectic" is exactly the phrase Kant used in the *Critique of Pure Reason* when speaking of the unavoidable tendency of theoretical pure reason to fall into illusions about things beyond its limits. Even after one learns the lessons of Transcendental Idealism, we still cannot help but to fall into this "natural dialectic" which pushes us toward illusions about metaphysics:

«[T]here is a natural and unavoidable dialect of pure reason, not one in which a bungler might be entangled through lack of acquaintance, or one that some sophist has artfully invented in order to confuse rational people, but *one that irremediably attaches to human reason, so that even after we have exposed the mirage it will still not cease to lead our reason on with false hopes, continually propelling it into momentary aberrations that always need to be removed*» (A298/B354, emphasis added).

A "natural dialectic" is, then, a sort of disease to which reason is prone, one that cannot so much be cured as managed via constant vigilance. I apply all of this to Kant's discussion of the "natural dialectic" of *practical* reason as well. We will never *not be* tempted to self-deception about the strictness of moral laws. Rather, we must maintain a vigilance which assures we will recognize when this natural tendency kicks in. This pervasiveness of the possibility of self-deception reaffirms again Kant's original point, viz., that *knowledge* of

oneself is the *first* command of all duties to oneself. This duty is as central as it is because of the nature of finite practical reason itself.<sup>4</sup>

Self-deception is *the* human failure with which one needs to come to terms in order to become moral. And self-knowledge is the first of all duties to oneself overall: if it is in “human nature” to engage in the sorts of deception we have analyzed here, then it is impossible to be a moral person (or even to understand the nature of the duties to which one is held) without removing that obstacle to morality that is the Dear, Deceiving, Devious Self.

## **Bibliography**

### **Kant’s writings**

All reference to Kant will first reference the Akademie edition page numbers, followed by the page numbers of the following Cambridge translation of Kant’s works:

*Critique of Pure Reason* (1999). Paul Guyer and Allen Wood, trs., Cambridge, Cambridge University Press.

*Groundwork of the Metaphysics of Morals* (1998). Mary Gregor, trs., Cambridge, Cambridge University Press.

*The Metaphysics of Morals* (1996). Mary Gregor, trs., Cambridge: Cambridge University Press.

### **Other references**

Austen, J. (2008). *Emma*, Oxford: Oxford University Press.

Grenberg, J. (2013). *Kant’s Defense of Common Moral Experience: A Phenomenological Account*. Cambridge, Cambridge University Press.

———(2005). *Kant and the Ethics of Humility: A Story of Dependence, Corruption and Virtue*. Cambridge, Cambridge University Press.



---

<sup>4</sup> It is this inextirpable tendency toward self-deception that inspires my phenomenological reading of Kant (see: Grenberg, Jeanine, *Kant’s Defense of Common Moral Experience: A Phenomenological Account* [Cambridge: Cambridge University Press, 2013]). Attentive phenomenological reflection is the antidote to this inextirpable tendency toward self-deception. Seeing well how moral reasons press themselves on us is the first step in becoming a moral agent. Choice is a second step that is accomplished most successfully by prefacing choice with this moral attentiveness.