Interview

Amélie Oksenberg Rorty *

Edited by Patrizia Pedrini

AMÉLIE OKSENBERG RORTY is the Findlay Professor of Philosophy at Boston University and a Lecturer in the Department of Global Health and Social Medicine at Harvard University. Her Mind in Action (1988) consists of essays in ethics and philosophical psychology; she has also published a number of anthologies on Aristotle’s ethics, his poetics and his rhetoric, as well as a collection of papers on Descartes’ Meditations. Continuing her interest in the philosophy of education (Philosophers on Education, 1998), she is now working on a book defending ambivalence: On the Other Hand: The Ethics of Ambivalence.

With Brian McLaughlin, Rorty edited and contributed to a seminal collection on self-deception (Perspectives on Self-Deception, 1988) which is still a classic on the subject. Rorty and McLaughlin acknowledge that «explaining, or explaining away, the phenomena of self-deception raises many of the central problems in the philosophy of mind» and rightly declare that they use «self-deception as a microcosmic case study that bears on a range of issues dividing contemporary philosophical psychology», because

[…] disagreements about the existence and analysis of self-deception expresses disagreements about the unity of consciousness, homuncularism in psychological explanations, the criteria for the attribution of belief, the conditions of intentionality and rationality, the primacy of cognition in psychological processes, the relation between motivational and epistemic attitudes, the social formation and malformation of belief and self-deception, and moral constraints on responsible belief. (McLaughlin & Rorty, 1988, p. 1)

Rorty and McLaughlin were aware of the importance of these topics for epistemology and ethics, as well as the philosophy of mind. They therefore divided the collection into sections covering “The Analysis of Self-Deception” (part I), “The Epistemic Dimension of Self-Deception” (part II), “The

* I am grateful to Patrizia Pedrini for her searching questions and to M.R. Amiran, Aaron Garrett, Steven Gerrard, Aryeh Kosman, William Ruddick, and Richard Schmitt for helpful comments
Psychology of Self-Deception” (Part III), “The Social Dimension of Self-Deception” (Part IV), “The Moral Dimension of Self-Deception” (Part V), and finally also “Self-Deception and Literature” (Part VI).

This interview tries to focus both on Professor Rorty’s explanation of self-deception and on her views on some ongoing open questions and recent controversies. I asked Professor Rorty to answer six questions, to which she offered extensive, challenging responses. We are all most grateful to Professor Rorty for having generously undertaken this task.

1. In your seminal work on self-deception, you defended the idea that self-deception becomes less mysterious once we accept a conception of the self as a «loosely organized system of relatively autonomous subsystems» (Rorty, 1988, p. 12). The view you held in the paper quoted was brilliantly capable of accommodating a phenomenon that Donald Davidson’s view was perhaps making unnecessarily paradoxical. In this sense, you anticipated the spirit of Al Mele’s “deflationary view” of self-deception (2001). Would you still subscribe to this view of the self and to how it applies to the explanation of self-deception, or have subsequent reflections changed your mind on this point, or refined your position?

I think that the familiar philosophical puzzles about the apparent incoherence of self-deception rest on views about the ‘the self’ as a unified and temporally continuous entity capable of acting from rationally monitored reflective self-awareness. So construed, the idea of the self is a theoretical construction, designed to accommodate cultural notions of individual agency and responsibility. Largely for the sake of rationalizing our practices of assigning responsibility, we treat the self as a psychologically and cognitively unified entity, capable of effective self-knowledge. The range of actions for which we hold individuals responsible varies with what we take to be within a normal agent’s knowledge and reflective capacities. On the one hand, we hold individuals morally and legally responsible for a wide scope of voluntary agency, including their intentions as well as their actions; on the other hand, we accept a wide and generous latitude of excusing conditions to explain and exonerate failures of responsibility.

I believe that the idea of the self as a unified, conscious and presumptively self-aware entity is an ideal superimposed on a loosely organized system of
relatively independent but mutually supportive and interactive modular psycho-physical subsystems, only some of which are capable of ‘internal’ scanning. As Carruthers puts it,

[such] modules might be isolable function-specific processing systems, all or almost all of which are domain specific, and whose operations aren’t subject to the will. [These modules] are associated with specific neural structures (albeit sometimes spatially dispersed ones). [Although these modules are typically interactive] [...] their internal operations may [sometimes become] [...] inaccessible to the remainder of cognition. (Carruthers, 2006)

Van Leeuwen goes further: «The capacity for self-deception [...] is a spandrel [...] of other mental traits, i.e., a structural byproduct. The irony is that the mental traits of which self-deception is a spandrel/byproduct are themselves rational» (van Leeuwen, 2007). 1 Although individual persons are presumed to be normally conscious, capable of basic reflective introspection, the scope of their capacities for accurate self-awareness varies considerably. For instance, some people have acute self-knowledge in epistemic matters, but very little understanding of their motivational patterns: they are good at reflecting on what they believe, but are often mistaken about what they desire. Others are sensitive to their sensory and proprioceptive functioning but relatively unreflective and often mistaken about what they value.

Some modular sub-systems of the self function as internal scanners, dispositionally geared to monitor cognitive and psycho-physical operations as the need arises. Individuals vary 1) in their ability to coordinate scanning information with other cognitive and connative functions and 2) in the extent to which they can voluntarily control and direct their scanning operations. Some areas of psychological and cognitive functioning — for instance, high order cognition engaged in theoretical reasoning — tend to be more transparent than those engaged in preferences that were developed in infancy. For some people, conflicts of beliefs and desires are relatively transparent, easy to diagnose. Although they may find such conflicts troublesome, such people may be less subject to self-deception than are those who resist or deflect reflective scrutiny of conflicting beliefs and desires. Patterns of accessible scanning and accurate reporting can be affected by trauma; self-knowledge can become more or less acute with experience and with motivational changes. The more integrated and voluntary are a person scanning functions, the less is she

1 See also Fodor 2000.
likely to be subject to self-deception. On the other hand, those with a low level of epistemic integration — those who tend not to monitor the consistency of their sub-systems — may simply be inconsistent or be mistaken about their beliefs. Because they never claimed self-knowledge, they may not be self-deceived. A great deal of apparent self-deception involves a contrast between the content of a conscious occurrent belief and that of an unacknowledged — and sometimes vague — dispositional belief. Because the criteria for the attribution of various types of belief vary, and because its ascription can be a matter of degrees, there may sometimes be more (and sometimes less) self-deception than meets the eye. In any case, self-deception is notoriously difficult to ascribe with any confidence because it typically occurs in opaque contexts.

2. The idea of the adaptive fitness of self-deception had been first and importantly defended in your writings on the topic. However, not all the scholars agree that all forms of self-deception are invariantly adaptive for the species, let alone that it will always make us flourish individually (according to criteria for the “flourishing” in question that a scholar might want to specify) or make us happy (e.g., Van Leewen, 2009). What’s your thought about the new arguments produced by those who are sceptic about the adaptive value of self-deception?

The structural capacities for self-deception — the relative independence and compartmentalization of psychological and cognitive sub-systems — are adaptive for survival and for high level functioning. The functional independence of such subsystems promotes specialized and highly developed cognitive and psychological activities; it enables intensive focused attention; it protects sub-systems from doing infectious collateral damage to one another; it enlarges the diversified scope of psychological and cognitive functioning. By bracketing agents’ awareness of risk, it enables them to act with confidence and conviction in situations of uncertainty and risk, to be devoted to personal and social commitments when closer scrutiny might distance them, to maintain an even tempo and temperament in the face of the erratic fluctuating circumstances.

2 See also discussions in Martin 2009.
To be sure, not every instance—or even every type—of self-deception is beneficial, either for the individual or for the species. The psycho-physical structures that are adaptive for effective psychological functioning nevertheless also bring marked disadvantages and vulnerabilities. Functionally and structurally independent sub-systems increase the possibility of the failure of psychological integration; they can conduce to the kind of active disintegration that self-deception and akrasia sometimes represent. The benefits of compartmentalized functionally independent sub-systems are matched by the need for their integration, for accurate transparency and accessibility among them. Because the effective strategies of psychological and cognitive adaptivity are integrally connected to their vulnerabilities (and vice versa), their integration requires constant adjustment in ways that are rarely under voluntary or even conscious control. Ironically, such adjustment obviously presupposes the very integration it is meant to maintain.

Given the advantages of the structural capacities for self deception and the benefits of a great deal of self-deception, why does it have such a bad press? Why do we blame ourselves and disdain others for what is in many ways an adaptive and useful strategy, one that sustains many of our central activities? At least one of the draw backs of self-deception is that it is a powerful instrument of moral indifference and even cruelty. Consider how a self-deceiver might deflect criticism of his behavior by describing a shady negotiation as resourceful rather than as aggressive or by describing a fawning and flowery compliment as tactful rather than hypocritical. The brilliantly inventive and self-deceptive ability to find or to concoct a covering but deflecting description for a morally suspect action can provide the basis for a tangential moralizing justification that masks and disguise great wrongs. It enables us to blind ourselves to our motives and to the effects of our actions on others; even more dramatically, it enables us to ignore or misdescribe what we are actually doing. Self-deception allows us to abstract ourselves from our actions, remaining selectively ignorant of their presuppositions and consequences. Kant’s severity describes the matter well: «[The] inner advocate expounds the law to [his] advantage [...] he grows deceitful, making use of the law for his own purposes, [as] a means of self-deception whereby he persuades himself that he has been acting rightly, on principle» (Kant, 1963, p. 137). Taking advantage of Kant’s emphasis on the freedom of self-legislation, the self-deceiving Mafioso within adds: “You want moral principles? I can get them for you wholesale.”
How does the apparently innocent self-deceiver manage to bring off his own deception? Self-deception is sometimes a free rider on referential opacity (Kaplan, 1986). Even the smallest, most precise actions or character traits are open to multiple descriptions whose tonal connotations, etymologies and classifications implicitly tend to direct its evaluation and justifiability. Although such descriptions are not substitutable salva moralitate, the louche self-deceiver treats them as fungible: she substitutes a morally permissible description of an action or trait for one that might be morally suspect. By treating a referentially opaque expression as if it were transparently substitutable salva moralitate, she gains ground for justifying the action to which it refers. Referential opacity allows the ingenious self-deceiver to find a resonant principle to justify whatever interests she favors by focusing on an astutely self-serving description of what she does. All she has to do is emphasize some features of her traits or actions as salient, others as recessive. Without actually lying to herself, the self-deceiver can present herself to herself as a morally decent if not actually estimable figure.

Hannah Arendt (2006) argued that the failure to think, the failure to notice or attend to the full description of what we do is often the first step in finding a convenient, apparently reasonable justification for great wrongs. Self-deception can take the form of astutely substituting a thin and morally innocent description of an action for one that would reveal its morally relevant thick description. Consider Eichmann defending himself by saying “I was just following orders to coordinate train schedules.” That thin generic description of his action carries relatively neutral implications and expectations about its generic standard aims, settings, and outcomes. It carries an implicit standard justificatory explanation that tends to deflect the kind of attentive questioning that might press for a fuller, thicker description. A more robustly detailed thick description – “I consulted train schedules to plan a timetable for transporting gypsies to Auschwitz” – might have unmasked Eichmann’s self-deceptive justification of what he did. But neither the thin nor the thick description of Eichmann’s scheduling trains to Auschwitz necessarily reveals his motivational structure: he might have been an ordinary standard issue bureaucrat, primarily focused on doing whatever would undermine his rival in the SS Schutzstaffel. Or he might have been an obsessive compulsive, a man with a tidy, obedient mind whose attention was always focused on the minutiae of whatever he did. Quite independently of his motives or habits, Eichmann can be self-deceived about (the thick description of) his action in constructing a schedule for
transports to Auschwitz. The brilliantly inventive and self-deceptive ability to find or to concoct a covering but deflecting, tangential moralizing justification can mask and disguise moral failures. Eichmann might — or might not — have been self-deceived about his motives as well as about his action. His being self-deceived about his motives might — or might not — have explained his being self-deceived about his action. In any case, the evaluation of his motives is independent of his being self-deceived about what he did.

3. A very new question raised by Eric Funkhouser (2005) is what the self-deceiver wants and whether it ultimately gets what he wants. The controversy is still live and attracts much interest, and I would like to ask you about your current view of the motivational state of the self-deceiver.

Sometimes self-deception just happens: a self-deceiver need not always be motivationally prompted to deceive himself about his beliefs or about anything else, for that matter. A pattern of self-deception can become habitual as a result of a person’s psychological history or his social milieu, without any particular motivation on his part. (Ruddick, 1988). Just as a painter can deceive a biographer or art historian, so too she can deceive herself about the merits of her work. Because her parents and friends successfully deceived her about her talent, she came to collude in the deceptive estimation of her talent. To be sure, sometimes such a painter may simply be chronically mistaken, but she might sometimes actively collude in keeping herself from realizing the truth of the matter. She can consistently be inventively obtuse, ignoring or denying the evidence given by critics, collectors and museum curators whom she normally admires and whose judgment she trusts. Her self-deceptive self-esteem can be habitual, without being specifically motivated.

In any case, not all self-deception is deception about the self, or about its beliefs and desires. Very roughly, X is self-deceived about p (where p can be any state of affairs) when 1) X has evidence that p, and 2) X directly or indirectly denies that she has evidence that p (or believes q, where X has evidence that q entails not-p); and 3) there is evidence that X is aware that she both believes and denies that p; and 4) X directly or indirectly denies that she has such evidence. In the second place, although these affirmations and denials can sometimes be motivated, they need not be prompted by a specific concurrent desire. To be sure, beliefs are, in the very nature of the case, truth-directed and truth-claiming, presumptively integrated in a truth-oriented
system of beliefs. In that sense, belief-claims carry second-order implications about the believer’s commitment to truth-orientation. Those commitments need not, however, indicate anything about his wants or desires. It is not unusual for someone to want to be free of his commitments: he might sometimes not altogether unreasonably wish he were less committed to telling – or even to discovering – the truth. ‘Beliefs’ that are fully constituted or determined by non-truth-tracking second order motivations are nevertheless suspect as instances of bona fide beliefs, independently of whether they are self-deceived. Expressions of wishes rather than of beliefs, they may prompt self-deceptive claims without themselves being instances of self-deception.

4. Self-deception seems to involve a failure of self-knowledge (e.g., Scott-Kakures, 2002). Do you think this is correct and how would you characterize this failure?

Self-deception does not involve more failure in self-knowledge than we ordinarily have under ‘normal’ circumstances. We have very little self-knowledge to begin with: we are rarely able to articulate the scope and details of our values and commitments; we are often mistaken about our basic character traits; we are often at sea about whether we are prepared to affirm the logical entailments or presuppositions of propositions we take ourselves to believe. The limitations of self-knowledge do not necessarily involve self-deception: they typically indicate ignorance, diagnostic errors and sometimes simple disinterest. On the other hand, since self-deception is not necessarily deception about the self, not all self-deception involves a failure of self-knowledge. Sometimes it involves denials in the face of overwhelming evidence of the chicanery of friends or the corruption of colleagues.

Just as deception does not necessarily involve lying, so self-deception does not necessarily involve holding a false belief. It is possible to mislead or deceive someone by distracting them, by redirecting their attention to some inane or trivial truth. So too one can deceive oneself by paying careful and accurate attention to some distracting or tangential feature of one’s experience, and so mislead or deceive oneself to ignore what might be most germane in the circumstances. As I suggested in my response to Question 2, referential opacity is the self-deceiver’s friend: sometimes the canny self-deceiver need only substitute an alternative description of an action a description that captures his focused attention – to deceive himself about what he is doing.
5. What’s your current view about the relationship between confabulation and self-deception? Hirste (2004) argues for the view that there are some overlaps between the two phenomena but the debate is still open.

Some — but by no means all - self-deception is accompanied by a covering confabulation designed to explain away awareness of counter-vailing evidence to cherished or entrenched beliefs. It seems to me that “the overlap view” is overly and nervously intellectualistic: in practice, in the ordinary course of things, neither believers nor deceivers feel the need to explain —or explain away — the grounds for their attitudes.\(^3\) Just as we do not confabulate to explain errors of judgment unless we are pressed to do so, so too we do not typically need to explain consistently deflected attention by confabulating. Indeed confabulation tends to highlight the self-deception, to make it suspect. Qui s’excuse, s’accuse. Self-deception typically remains unacknowledged and unexplained: the entrenched self-deceiver standardly overlooks the pattern of his denials. Of course someone charged with self-deception — given solid evidence of its occurrence — sometimes confabulates to explain or exculpate himself. In such cases, confabulation accompanies self-deception without being integral to its strategies.

6. Finally, do you think there is any urgent question scholars should address in order to make the current research on self-deception progress further in the light of the new results in philosophy of mind?

- Our understanding of self-deception would benefit greatly from research into the structures of localized, modular sub-systematic patterns of brain functioning and from studies of the integration of cognitive centers with endocrine functioning. Under what conditions does such integration succeed and when does it fail?
- Inter-disciplinary studies in the philosophy of language and the psychology of speech acts — analyses of the relation between the psychology of propositional attitudes and the pragmatics of speech acts — would also be illuminating. What kinds of speech acts qualify as self-deceptive? Can merely expressive non-propositional utterances be self-deceptive? Can wishes and fantasies be self-deceptive? What is the

\(^3\) See the classic studies reported by Richard Nisbett and Lee Ross (1980) and by Daniel Kahneman, Paul Slovic and Amos Tversky (1982).
structure of self-deceptive promising? Can performative or constative speech acts be self-deceptive? Can externalist and internalist standards of the attribution of self-deception be reconciled?  

- Anthropological and sociological studies of self-deception would enlarge and correct our present rather provincial understanding of the dynamics — and the norms — of self-deception. Do cultures differ in the domains in which self-deception is prevalent? What sorts of social pressures support or conduce to self-deception? Does the prevalence of forms of high politeness in social inter-action conduce to self-deception? What are the cultural differences in the incidence and areas of common self-deception? Does successful self-deception typically involve social reinforcement? Can religious or social rituals like absolution, forgiveness, penitential prayers be self-deceptive?

- Victorian novels (George Eliot, Trollope, D’Israeli) and political autobiographies (Koestler, de Beauvoir) provide wonderful insight into the subtle processes of self-deception and their occasional unmasking. We have, for instance, much to learn from tracing Eliot’s descriptions of Dorothea’s self-deceptive admiration for Casaubon and her gradual, reluctant disillusionment. Lydgate’s blindness to Rosamond’s manipulations highlights the way that naïve self-deceivers sometimes collude in the deceptions that others initiate. Autobiographies of fervent communists who became anti-communists after the Stalin Trials also provide rich examples of the reflections of self-declared self-deceivers, of the strategies they employed in their self-deceptions, of their techniques in resisting contrary evidence, of the occasions of their “breakthrough” self-corrections.

- The current industry of philosophical work on self-knowledge — initially prompted by Anscombe (1981) and recently developed by Holton (2009), Bermúdez (1998), Cassam (1994), Gertler (2003), Moran (2001), and Hatzimoysis (2011) — would benefit from a closer study of the various domains and strategies of self-deception. It would also be illuminating to locate the varieties of strategies of self-deception within a

---

5 For novels, see George Eliot, Middlemarch, 1874; Trollope, He Knew He was Right, 1869; D’Israeli, Sybil, 1845; Henry James, The Wings of the Dove (1902); for autobiographies, see Arthur Koestler, The Invisible Writing and The God that Failed (1949), Simone de Beauvoir, The Force of Circumstance (1963) and All Said and Done (1972).
more general taxonomic frame of the varieties and domains of belief and of self-knowledge.

- Is the idea of collective or interactive self-deception coherent? Can analyses of collective intention and action be applied to self-deception? (Gilbert, 1989; Bratman, 2007). If so, how does it work, what are its ‘mechanisms?’ What are its implications for the philosophy of mind and the philosophy of language?

- We need a catalogue and taxonomy of the varieties of self-deception, with an account of how their domains and strategies differ from the varieties of self-knowledge. Is there a significant difference between motivated and non-motivated self-deception, between its occurrent and habitual forms? between self-deceptive belief and self-deceptive action or emotion? between direct or active and indirect, passive or collusive self-deception? between self-deception that issues in false belief and that which issues in a true but pragmatically defective belief or action?

I am especially interested in indirect, passive or collusive self-deception, cases where we collude in being deceived by others. Consider the ways that we knowingly allow ourselves to be conned, “taken in” by political rhetoric and manipulative advertising. We typically know perfectly well that such claims and promises are inflated if not actually false, and yet we find ourselves believing and acting as if they were reliable and trustworthy. What makes us susceptible to internalizing claims that we would typically hold suspect? When and why do we abandon our normal epistemic caution and extend epistemic trust beyond its normal limits?

- I suspect that self-deception is now a fashionable topic in the philosophy of mind because a great deal of post-Wittgensteinian philosophical psychology has focused on perceptual and cognitive transparency.⁶ The prevalence of philosophical concern about self-deception is also of concern to consequentialists and neo-Kantians who place heavy emphasis on the underlying unity and effectiveness of the capacities for rational choice or self-construction.⁷ Chronic and structural vulnerability to self-deception — endemic and apparently functional patterns of irrationality — appear to threaten effective norms of the rational basis and directives of

---

⁶ See e.g., Siegel 2010.
⁷ See e.g., Railton 2003 and Korsgaard 2009.
morality. Integration, self-knowledge and integrity are in high demand precisely because they seem elusive. We are concerned to eradicate self-deception because it seems to threaten our claims to epistemic responsibility, moral integrity and social reliability. The independence of modular sub-systems engaged in high level cognitive thinking from those engaged in sensation and perception — the focused reflexive and transparent awareness of abstract thoughts abstracted from perceptual content — is ironically the very condition that makes us capable of — and vulnerable to — self-deception and akrasia as well as other common and prevalent forms of irrationality.

REFERENCES


