1 Introduction

> For without philosophy, nowadays only criminals dare to hurt other humans.
> (Robert Musil, *Der Mann ohne Eigenschaften*, ch. 48; author's translation)

> L'hypocrisie est un hommage que le vice rend à la vertu.
> (François de la Rochefoucauld, *Réflexions ou sentences et maximes morales et réflexions diverses*, Maxim 218)

> So convenient it is to be a reasonable creature, since it enables one to find or make a reason for everything one has a mind to do.
> (Benjamin Franklin, *The Life of Benjamin Franklin: An Autobiographical Manuscript*, p. 43)

> The key thing with resolutions is not to keep them. It's how to revise them once you fail.
> (John Oliver on New Year's resolutions)

This Element combines two of my long-standing philosophical interests: the ethics of Immanuel Kant and misuses of rational capacities. The significance of the latter phenomenon stretches far beyond Kant scholarship. For instance, one of the most memorable moments of my student days was when one of my professors claimed that 'Ethicists are the worst people in the world because they know all the excuses'. I think this is a somewhat pessimistic – though, as we will see, very Kantian – thought.

According to Immanuel Kant's biographer Manfred Kuehn (2001: 222), Kant 'had formulated the maxim for himself that he would smoke only one pipe [a day], but it is reported that the bowls of his pipes increased considerably in size as the years went on'. Kant's implementation of his own maxim is here in tension with the aim incorporated into this maxim, which presumably was to limit tobacco consumption. Kant treats the maxim as an externally imposed constraint that must be obeyed to the letter, but not in spirit. The implementation undermines his end without formally renouncing it. Kuehn's anecdote, though it concerns matters of prudence rather than morality, nicely illustrates the subject matter of this Element: quibbling with rules, trying to outsmart one's better self or, as Kant calls it, *vernünfteln*/'rationalizing', understood as the use of rational capacities to undermine reason, or an exercise of reason that weakens agents' readiness to do the right thing while they yet deem themselves committed to morality.

Kant was a keen psychological observer of this phenomenon, and the dangers it poses are a major theme in his moral philosophy. Consequently, his discussion of rationalizing will help us to better understand a number of important aspects of his philosophy. First, practical philosophy is supposed to function as an

'antidote' (Wood 2002: 28) against rationalizing. Understanding rationalizing
will illuminate the *purposes that moral principles and theories serve*, and will
help us to understand what the role of the philosopher is for moral
improvement.[1] Second, the concept of rationalizing is central for our under-
standing of Kant's engagement with his *academic colleagues and the popular
philosophers* whom he considers to advocate sophisticated forms of rationaliz-
ing. Moreover, he criticizes religious institutions and practices because they
propagate and encourage mistaken beliefs about moral responsibility.[2] Third,
Kant's discussions of concrete examples of rationalizing offer instructive *case
studies for how our moral reasoning can go wrong*. A detailed look at these
examples will enhance our understanding of Kant's conception of the workings
as well as deficits of our rational faculties. Fourth, throughout his discussions of
the various *dialectics* that our reason is subject to and that necessitate critiques
of our rational capacities, Kant describes fundamental mistakes in reasoning
labelled 'rationalizing'.[3] Kant is interested in the impact of rationalizing on all
aspects of our use of reason. However, due to constraints of space, I will focus
on the practical dimension.

   In this Element, I will take a detailed look at the examples Kant provides of
rationalizing, at his own explanations of the underlying process and at the

---

[1] Wood (2017: 20–6, 74) correctly stresses that Kant's ethics is not primarily supposed to serve
a theoretical or intellectual purpose, but to address *moral* flaws. Most recently Callanan (2019)
has argued that the discussion of the natural dialectic, in which Kant introduces rationalizing, is
central to a correct understanding of the *Groundwork*. Kant here parts way with Rousseau who is
sceptical that philosophy can ever help agents to become and remain moral, whereas Kant argues
that philosophy can and must perform this function.

[2] We can find a secular example for this phenomenon in a *New Yorker* cartoon in which a number of
men in a business meeting ask their secretary, Miss Dugan, to send in an expert who can tell right
from wrong. Wood (2017: 17), from whom I take this example, analyzes this as follows. These
men 'are about to do something they know is wrong. Yet they are tempted to do it anyway, no
doubt on the ground that doing it serves "the greater good" (the firm's, the university's, or just
their own). They are in a quandary because they are tempted to think that this "greater good"
might justify (perhaps only "just this once") their doing what [. . .] they know perfectly well is
wrong. The call to Miss Dugan is an admission that, in their condition of moral weakness, the
shallow "cognitive" (i.e., the "cost–benefit" or "greater good") part of their brains has so
disoriented their good judgment that they no longer know what they know and what they don't.
But at least they do know that they no longer know what they know; that last pitiful shred of
human decency shows itself in their desperate plea for help, comically masquerading as
a dignified professional request for outside expertise.' There are many other contemporary
examples from politics, public discourse, and the private sector as well as institutions of education
to which Kant's analysis of rationalizing applies.

[3] This is most apparent for the natural dialectic in G, IV: 404.37–405.19, which I will discuss in
detail in Section 3. In the First Critique, rationalizing is presented as an exercise of rational
capacities without awareness of their dialectical nature, resulting in antinomical claims (A/B:
422/450; see also A/B: 63/87–8, 421/448–9). In the Third Critique, Kant explains rationalizing as
the act of claiming a priori universality for one's judgements, which can lead to a dialectic of
opposing judgements (CJ, V: 337.5–8; see Section 2.1).

effects rationalizing has on agents' grasp of morality, and will discuss the
necessary conditions for rationalizing to be possible and to appear as
a promising strategy to agents for dealing with moral commands that can require
great sacrifice from them. In doing so, I aim to shed new light on Kant's
philosophy in a number of ways.

(1) The main contribution of this Element is to challenge the widely shared
assumption that rationalizing in Kant only extends to *questions of motivation
and specific maxims*. By drawing on the full breadth of examples of rationaliz-
ing presented by Kant, I will show how rationalizing can lead to a systematically
distorted sense of right and wrong which Kant labels 'corruption'. When
I rationalize, I undermine the grasp that the supreme principle of morality has
on me, while I still believe myself fully committed to morality, albeit to a less
demanding moral principle or more lenient conception of morality. There is
much more to rationalizing than misrepresenting the strength or source of
motives.

(2) I explain in what sense rationalizing is a *rational* activity. Empirical
practical reason devises pseudo-justifications and finds excuses to promote
an agent's sensuous ends at the expense of morality. Paradoxically, the interest
in these pseudo-justifications is rooted in rational agents' recognition of the
authority of morality. Only agents who recognize the authority of morality are
tempted to look for excuses or apparent justifications for their morally dubious
actions. My discussion will reveal that Kant is not an arch-rationalist who
believes in the power of reason without qualification.[4] Kant understands that
moral failings are not simply the fault of inclinations and of our sensuous side
(in fact, they are never simply that). Many instances of moral failure are
expressions of fallacious (but not always obviously incorrect) reasoning and
even of forms of pseudo-rationality that can be extremely sophisticated. Kant
is aware that rationality is a double-edged sword; it is the source of morality
and of our dignity, but it also enables us to seemingly justify moral transgres-
sions to ourselves and others, and it creates an interest in such justifications in
the first place.

(3) I explain how it is possible for a rationalizer to think that committing
a moral violation can be *excused* and even *justified*, even though this rational-
izer has not completely lost touch with the moral law. This will allow us to

---

[4] That Kant was blindly optimistic about the power of reason is a stereotype that still prevails at
least among non-specialists. See, for instance, Haidt (2001), who discusses Kant under the label
'Worship of Reason', and Mercier and Sperber (2018: 17), who count Kant as a philosopher who
assumes that 'humans err by not reasoning enough', not by 'reasoning too much'.

maintain that rationalizers are still moral agents and are morally responsible for their actions. It will also demonstrate that Kant does not think that corrupted agents are merely in a state of confusion in which they feel a need for urgent philosophical help. They can be in a state of false (though never complete) certainty.

(4) I will close with a *critical discussion* of the scope and underlying assumptions of Kant's conception of rationalizing. Such a critical discussion is pivotal, since Kant criticizes competing ethical theories for being rationalizations and for reinforcing and encouraging rationalizing. Understanding whether charging a philosopher with rationalizing is a valid criticism, and understanding whether Kant is in a position to level this criticism against other theorists, will help us gain a better general understanding of the strengths and weaknesses of Kant's ethics.

2 Rationalizing in Context

I begin with a few remarks about the term 'rationalizing' (Section 2.1) and about common themes in the literature on rationalizing and self-deception (Section 2.2); I then outline why rationalizing is an important concern specifically in the context of Kant's *practical* philosophy (Section 2.3).

2.1 *Vernünfteln*

The nouns *Vernünfteln, Vernünftelei* and *Vernünftelung*, as well as the verb *vernünfteln*, were in much more frequent use in the eighteenth century than they are in contemporary German, where the term is largely out of use. *Vernünfteln* is composed of *Vernunft* (reason), and the suffix *eln*. This suffix indicates that *Vernünfteln* is an activity. It is also a diminutive (see Adelung 1811: vol. 3, col. 1785), which gives *Vernünfteln* negative connotations in the sense of using reason in a deficient manner. According to Grimm and Grimm (1889: vol. 25, col. 936–9), *Vernünfteley* (noun) is an 'incorrect and inappropriate use of reason', and *vernünfteln* (verb) means 'presenting something as seemingly correct, which is in itself pedantic and incorrect' (my translations).

Kant himself explains the meaning of the term twice. According to the Third Critique, *Vernünfteln* is the act of claiming a priori universality for one's judgements, which can lead to a dialectic of opposing judgements (CJ, V: 337.5–8): 'A rationalizing judgment (*iudicium ratiocinans*) is any judgment that declares itself to be universal' (CJ, V: 337fn.). *Vernünfteln* here is a necessary but not sufficient condition for dialectical reasoning. Only judgements claiming a priori universality can give rise to a dialectic. However, many

judgements claiming a priori universality do not lead to a dialectic.[5] After all, their claims might be justified. *Vernünfteln* here seems to not be inherently problematic. By contrast, in the *Anthropology*, 'Vernünfteln (without sound reason)' is defined as 'use of reason that misses its final goal, partly from inability, partly from an inappropriate viewpoint' (Anth, VII: 200.5–7). *Vernünfteln*, if not constrained by sound reason, is here presented as a use of reason that, by definition, produces incorrect results. Moreover, *Vernünfteln* is presented as of a much broader scope, since it is not necessarily tied to claims to (a priori) universality and it can result in all kinds of mistakes, not all of which have to be dialectical.

Kant's actual use of the term is even broader than what the *Anthropology*'s definition indicates. He sometimes uses the term to simply mean 'reasoning' or 'reflection' without any negative connotations. This is presumably the case when *Vernünfteln* is constrained by, or in the service of, sound reason.[6] In many cases, *Vernünfteln* does, however, indicate a *deficient* use of rational capacities,[7] and I will discuss in Section 3.1 whether this deficiency involves a dialectics in the terminological sense. Furthermore, Kant sometimes uses other terms to mean the same as *Vernünfteln*, occasionally using *vernünfteln* and *räsonniren* interchangeably (Anth, VII: 200.10),[8] and he stresses that humans 'artificialize' [*künsteln*] in order to represent moral transgressions as an 'unintentional fault'

---

[5] Cf. also Guyer and Matthews (2000: 213): a rationalizing judgement connotes 'only a necessary condition of a sophistical argument, namely that it make a pretense to universality, without yet implying that anything that gives rise to a dialectic is sophistical in the usual, pejorative sense'. Kant distinguishes between a rationalizing judgement and a rational judgement [*Vernunfturteil*] (CJ, V: 337fn.). Only the latter is grounded a priori as it is the conclusion of a rational inference, whereas a rationalized judgement is the major premise of such an inference. Kant also frequently distinguishes 'rationalizing' from 'rational' (A/B, 340/398, 604/632; CJ, V: 396.7–16; RO, XX: 179.12–15).

[6] There is no German word equivalent to the English 'reasoning' or 'reasoner' in the sense that it is a cognate of 'reason'/*Vernunft* and indicates the use someone makes of reason or the person who reasons. Obviously, there are terms for this (e.g., *denken* and *Denker*) but these terms are not etymologically connected to reason. The closest German comes to such terms is *vernünfteln*/ 'rationalizing' and *Vernünftler*/'rationalizer'. Kant sometimes uses these terms to simply refer to 'reasoning' and 'reasoners'. I am grateful to Thomas Sturm for discussion of these etymological points.

[7] For examples of deficient uses of reason characterized as 'rationalizing', see A/B: 269/325–6; MFNS, IV: 505.25–31, 523.25–9; CJ, V: 262.24–5; MM, VI: 215.36–216.6; Anth, VII: 119.14–22, 201.12–23, 228.16–18; Men, VIII: 151.19–25; MPT, VIII: 267.18–22; TP, VIII: 301.1–15; TPP, VIII: 382.15–21; PCT, VIII: 416.3–9; Corr, X: 432.4–11; FI, XX: 234.12–24; PM, XX: 343.10–14; RE, XX: 412.13–413.11; P-PR, XX: 437.1–438.3; L-Men, XXV: 1041.28–1042.3; Eth-K: 127–9.

[8] However, often *räsonniren* is used to describe harmless arguments during dinner (Anth, VII: 280.10–19) or reasoning without sufficient insight into underlying principles (Anth, VII: 200.10–12, 476.31–4). Likewise, 'empty musing' [*leere Grübelei*] (Anth, VII: 221.30) and 'cheeky musings' [*vorwitzige Grübeleien*] (CF, VII: 24.2) indicate comparatively harmless shortcomings.

or mere 'oversight' (CPrR, V: 98.13–21)[9] and that they can 'bully' [*chicanie-ren*] their conscience (G, IV: 404.21; see also Eth-V, XXVII: 620.3).[10] Moreover, Kant refers to a vain or pointless exercise of rational capacities as *herausklügeln* (Rel, VI: 26fn.), which literally means 'to prudentialize something out of something'. Kant, while not usually known for his literary style, clearly shows great enthusiasm for the creative use of language when describing misuses of reason.

Vernünfteln has been translated in various ways.[11] I will follow Timmermann (2011) and use 'rationalizing', as it makes clear that *Vernünfteln* is an exercise of rational capacities. 'Rationalizing', as I will use the term, is self-deception about *moral* matters. There are other kinds of self-deception, such as about what is prudent, one's capabilities and social standing and maybe even about purely theoretical questions. Self-deception about these issues, however, requires a framework different from the one that explains rationalizing in my sense, since these other forms of self-deception are not driven by a rational interest in being morally justified (see Section 4.3).

I should note that 'rationalizing', as I will use the term, is broader than the rational-tail-wagging-the-emotional-dog manoeuvre of finding post hoc justifications for decisions driven by emotions.[12] That rationalizing is post hoc and preceded by judgement is a frequent assumption especially in psychological approaches to the phenomenon (see, for instance, Cushman 2020: 1).[13] Rationalizing in Kant is not merely a matter of finding post hoc justifications. Misconceptions about morality can inform ex ante reasoning and can influence judgements and overturn agents' decisions – but not change them for the better (see Sections 4.5 and 5.2). Moral judgements, for Kant, are never merely driven by emotions, since a judgement is always a cognitive act. Yet, moral judgements can be incorrect and rationalized. Moreover, agents rationalize not merely to feel better about their judgements or to make their own actions comprehensible

---

[9] According to the *Kaehler Notes*, 'artificializing' [*künsteln*] is supposed to modify one's representation of morality 'until it conforms to the inclinations and leisureliness' (Eth-K: 356).

[10] I owe the references to *chicanieren* to Di Giuolio (2020: 277–8). Timmermann's (2011: 36) translation, 'engage in legalistic quibbles', makes the connection to rationalizing more apparent but is somewhat free.

[11] For instance: 'quibbling' (Timmermann 2007: 48), 'subtle reasoning' (Gregor 1996: 139), 'sophistical' (Gregor 1996: 649; see also Guyer and Wood 1998: 629), 'sophistry' (Wood and Di Giovanni 1998: 88), 'subtle argument' (Wood and Di Giovanni 1998: 99), 'ratiocination' (Wood and Di Giovanni 1998: 144), 'rationalistic' (Guyer and Matthews 2000: 213), 'reasoning' (Guyer and Matthews 2000: 371), 'pseudorationality' (Kemp Smith 1933: 327).

[12] See Haidt's (2001) influential article in which he argues that most of our moral judgements are driven by emotions, and that justifications are only obtained post hoc.

[13] However, Ellis and Schwitzgebel (2020: 23) note that '*rationalization in the pejorative sense*', that is, when it leads to morally criticizable behaviour, is typically prior to a moral judgement and seeks to vindicate the option already favoured by the agent (see also Sievers 2020).

to themselves, but because they genuinely do acknowledge the authority of reason (Section 4.3).

2.2 Rationalizing in Kant: The State of the Art

Kant's frequent use of the term *Vernünfteln*[14] is in stark contrast to the neglect this term has received in the literature until recently. Caygill's (1995) and Holzhey and Mudroch's (2005) Kant dictionaries lack any entry on *Vernünfteln*, Eisler (1930) only quotes a single passage (namely, CJ, V: 337) and König's article on *Vernünfteln* in the recent *Kant-Lexikon* (Willaschek et al. 2015) does not even mention Kant's practical philosophy.[15]

More recently, however, there has been increased interest in rationalizing, and a number of publications have presented sophisticated textual and systematic reconstructions of rationalizing in Kant and highlighted the significance of this concept for Kant's philosophy. Among these publications, three approaches are especially noteworthy for my own.

First, Laura Papish (2018: ch. 3) stresses that rationalizing centrally involves phenomena such as quibbling with things for which we have overwhelming evidence (74–9). For her, rationalizing is a violation of *epistemic norms* and it primarily impacts questions of justification.[16] I agree with Papish's focus on matters of justification and, in fact, one of my main quarrels especially with older literature on the topic (see below in this section) is the widespread

---

[14] The search engine of the online *Kant-Korpus*, which contains the first twenty-three volumes of the *Academy Edition*, gives thirty-five hits for *Vernünfteln* (noun), twenty for *Vernünfteley/ Vernünftelei* (noun), eighteen for *Vernünftler* (noun), seventy for *vernünfteln* (verb) and forty for *vernünftelnd* (adjective). This does not include exotic compounds such as *herausvernünfteln* (CPrR, V: 31.26; OAD, VIII: 220.33; MPT, VIII: 264.12), *wegvernünfteln* (G, IV: 456.2; CPrR, V: 154.3; PM, XX: 279.5) and *übervernünfteln* (MPT, VIII: 265f.26). In comparison, *factum* [*Faktum*] in the sense of 'fact of reason', a central term for Kant's ethics from the Second Critique onwards, only occurs eighteen times in its terminological sense (CPrR, V: 6, 31, 32, 42, 43, 47, 55, 91, 104; MM, VI: 252; Corr, XI: 58, 340; Corr, XII: 21; Ref, XIX: 612; OP, XXI: 21, 25, 36). Of the fifty-five quotes that Grimm and Grimm's (1889: vol. 25, col. 936–9) entry on *Vernünfteln* provides, thirty-three are from Kant's works alone. Kant was evidently particularly fond of this term even when the term was still in wider use than today.

[15] Most recently, Callanan (2019: 4) stressed that there 'has been surprisingly little analysis of the mechanism of the natural dialectic', rationalizing among them.

[16] Allais (2021: 49) also stresses epistemic failings, such as 'screening off' certain considerations. Of course, violations of epistemic norms can have important practical implications. Papish (2018: 110) argues that 'devotion to self-love becomes entrenched insofar as self-deception enables self-love to stake out new territory that it did not previously have and that outstrips our initial commitment to securing more banal and immediate objects of desire'. See also Wehofsits (2020), who argues that rationalizing can increase the hold of passions on an agent by revaluating and elevating their normative status, which 'leads to more comprehensive cognitive distortions' (14). The real danger of rationalizing is not that it seemingly licenses one-off transgressions, but its broader impact on agents' reasoning and character (see also Section 5).

assumption that rationalizing is merely or mainly a matter of misinterpreting one's motivation rather than of judging incorrectly about what one ought to do.

Moreover, Papish claims that it is one of the 'most important features of Kant's approach to self-deception [. . .] that reason is constitutively incapable of' rationalizing such that a rationalizer forms 'a particular belief according to which laxness and impurity are permitted' (2018: 84). Rather, laxness and impurity enter indirectly 'by means of the diversion of attention toward some pleasing alternative cognition' (2018: 84). In contrast to Papish, I do think that it is possible that rationalizers explicitly endorse a lax and impure conception of morality. In fact, I believe that this is what Kant suspects has happened to many of his academic colleagues who postulate happiness as the source of morality. I will present a theory of rationalizing that seeks to fully account for the potentially devastating impact that this activity can have on our understanding of morality and that can explain how agents can come to confidently assert overly lenient conceptions of morality and to even defend them against reasonable criticism.

I suspect that the differences between Papish's and my take on rationalizing are due to our respective focus on different paradigms of rationalizing. Papish approaches the phenomenon from relatively everyday cases that we find in the political and social sphere.[17] By contrast, I focus chiefly on more extreme cases that we can find in Kant's discussion of eudaemonists and of religious practices. In these cases, it becomes apparent that Kant does allow for the possibility of very corrupted agents who are more than merely confused about morality. My reading, I believe, can make better sense of the way Kant seeks to address rationalizing, namely via his ethical theory. If rationalizing was primarily a violation of epistemic requirements, then we would expect Kant to present guides to good *thinking* in order to combat this failing. However, Kant thinks that specifically 'practical grounds' compel us 'to take a step into the field of a *practical philosophy*' (G, IV: 405.22–4). We need *ethical* theory to address rationalizing, because rationalizing, in its most dangerous form, involves false beliefs specifically about morality and lack of commitment to the correct moral principles. Having said this, however, I should note that I regard Papish's and my approach largely as complementary in the sense that we focus on different paradigmatic cases of the same phenomenon. Moreover, Papish's discussion of

---

[17] Papish's (2018: 74–5) chief examples from Kant are the *political moralist* who brings up supposed empirical evidence against the (non-empirical) proposition that people should unite into a just state (TPP, VIII: 378), shifting to *irrelevant questions* in order to distract from the real issues (MM, VI: 318) and *switching between roles* as a private and official person (WIE, VIII: 37).

the epistemic dimensions of rationalizing constitutes a major contribution to our understanding of the phenomenon.

Second, Jeanine Grenberg (2013) presents an innovative new reading of the methods and aims of Kant's ethics, which she claims is centred on the first-person perspective of an agent experiencing respect for the moral law. As part of her investigation into the perspective of the agent, Grenberg extensively discusses rationalizing and corruption as the main threats to an agent's grasp of morality. One of her central claims is that even the corrupt person must still retain '*some* sense of the superior authority of moral principles, at least enough to realize she needs help maintaining that authority' (2013: 93). I agree with her that rationalizing would not make sense for an agent who does not recognize the authority of morality, and in Section 4 (especially 4.3) I will explain why this is so. However, I will also argue that while corrupted agents can be brought to doubt their mistaken views, Kant is not committed to the claim that all rationalizers acknowledge or feel the need for philosophical help. Rationalizing is dangerous because it can result in *false certainty*. Understanding this false certainty will help us understand why external help is important for agents, and the kind of help required.

Third, in a paper on the objection of 'moral overdemandingness', Marcel van Ackeren and Martin Sticker (2014: sec. 3) propose that Kant would consider it 'a model case of a rationalizing attack on our ordinary understanding of morality' that 'some contemporary philosophers consider high demandingness of a theory as something that calls for revision of this theory' (86).[18] They suggest that Kant could respond to an overdemandingness objection by showing how those who level the objection are, in fact, engaged in rationalizing. In Section 6, I will critically discuss whether rationalizing can indeed function as a response to objections of overdemandingness and as a criticism of lenient moral theories.

In addition to the current interest specifically in the concept of rationalizing, there is also ongoing discussion of this phenomenon under the more general label of *self-deception*. Literature on this topic has done much to stress the significance of this phenomenon particularly for Kant's framework,[19] especially with regard to self-deception about motives. In fact, it is a widespread assumption that self-deception in Kant is a matter of misrepresenting the motivation of one's actions. For instance, Nelson Potter (2002: 386) points out the 'centrality

---

[18] A similar charge, without appeal to Kant, was levelled by Wilson (1993: 278), who criticizes the overdemandingness objection as 'an ideology of academicians who are now, in a way they have never been before, part of a materially favored class'.

[19] See, for instance, Piper (1988: 298), who states that for Kant 'the really pressing motivational problem for actual moral agents is not *akrasia*, but rather self-deception', and Grenberg (2010: 162), who argues that an ethical approach, which deems vicious acts to be free and rational, 'requires a moment of rationalization, or even self-deception, in order to work'.

and overwhelming significance of self-deception' as 'a subterranean theme in
Kant's moral philosophy', where self-deception takes the form of either deceiv-
ing oneself about the *strength* of moral motives (387) or as 'the self-deception
by which we tell ourselves that we are acting for high and purely moral motives,
when our real motives relate to self-love' (388).[20]

Associating self-deception with misrepresenting one's motives, of course,
makes a lot of sense in the context of a philosophy that both stresses the
importance of the right motivation for moral worth and maintains the ultimate
opacity of our motives (see Section 4.3). However, I think it is very significant
that rationalizing does not merely pertain to presenting oneself as having acted
for high and mighty reasons. Since Kant's theory of rationalizing is scattered
throughout his works, it is easy to miss the full extent to which he analyzes
different forms of rationalizing, as well as its moral, psychological and social
dimensions. That rationalizing, besides motivation, affects an agent's reasoning
and general understanding of morality is pivotal for Kant, as rationalizing is
supposed to function as an explanation for how it is possible that rational agents,
capable of moral cognition, can have false beliefs about morality, think that their
moral violations are justified and even advocate mistaken moral theories.

Finally, I should mention that it is telling that Kant himself does not discuss
a rather obvious strategy to seemingly get around moral commands: phrasing
one's maxims in such broad or specific ways that they always pass the univer-
salization test.[21] I do not think that Kant means to deny this phenomenon. He is,

---

[20] Likewise, Darwall (1988) emphasizes the importance of the notion of self-deception for any
philosophy that focuses on motives and dispositions as opposed to consequences of actions, and
his discussion of self-deception focuses on Kant's conception of conscience and opacity of one's
own motives. Hill also understands self-deception as a misrepresentation of one's motives or as
assuming 'that our motives are good, whatever they are' (Hill 2012: 354). Likewise, Ware (2009)
discusses self-deception in the context of opacity of motives. In his recent book he stresses that
the 'tendency to deceive ourselves' is the tendency 'to construe our intentions in a flattering or
praiseworthy light' (Ware 2021: 15), but also concedes that rationalizing can issue in a 'genuine
error of deliberation' (26). Indeed, the focus of the debate has recently shifted somewhat away
from motives. Moeller (2020: 104), for instance, describes the attempt of deceiving one's own
conscience thusly: the internal 'defence advocate might interpret the incentive as mere negli-
gence rather than intentional wrongdoing'. Moeller does not reduce self-deception to the
question of whether actions are morally worthy. She is aware that an investigation of my
incentives can impact the question of whether what I did should count as a free and intentional
action and can be (fully) imputed to me. I think we can and should go even further than this;
attempts to excuse myself or advance spurious justifications for my actions do not have to draw
on matters of incentives at all. Finally, Allais (2021: 49) acknowledges the pervasiveness of self-
deception on Kant's framework in the sense that self-deception does not only pertain to
'particular actions or particular moral requirements in specific circumstances' but rather 'to
the rightfulness of our ways of live'. For this purpose, questions about motivation are important
but are not the only ones that matter; self-deception takes the form of presenting 'our motives and
ourselves to ourselves as better than they are' (46).

[21] That self-deception only concerns the permissibility of specific maxims is, for instance, main-
tained by Broadie and Pybus (1982).