**Do good people love themselves? On rational self-love in Kant[[1]](#footnote-1)**

[Forthcoming in *Kant-Studien*. Penultimate draft. Please cite only published version.]

**ABSTRACT**

Kant is frequently read as saying that all self-love is bad and that the virtuous agent is one who suppresses self-love as much as possible. This paper argues that this is mistaken and that the right kind of self-love – what Kant calls *rational* self-love – plays an important role in a successful moral life. It shows how Kant provides a detailed taxonomy of different kinds of self-love. He contrasts the (practical) incentive of self-love with the (pathological) feeling of it; self-love of benevolence with self-love of delight; and self-absorbed/selfish with rational/moral varieties of each. The paper then argues that, while the *Critique of Practical Reason* only identifies a self-absorbed variety of self-love of delight, self-conceit, it gains a rational counterpart in *Religion within the Bounds of Mere Reason*: “self-contentment.” This is a positive self-love of delight uniquely felt by the morally good person. It is suggested that this shift reflects Kant’s increasing appreciation for the affective dimension of virtuous life: for imperfect human beings the moral law must not only be worth obeying, but worth loving. Thus, while for morally bad agents, self-love and morality inevitably conflict, good agents can and should love themselves.

**KEYWORDS**

Love, self-love, self-contentment, virtue, feeling

In his moral theory, Kant[[2]](#footnote-2) identifies a supreme principle of morality that is intended to capture, with philosophical rigour, our intuitive grasp of some commonsense moral views. Prominent among them is that selfish actions are bad, whereas acting morally requires taking a perspective that goes beyond oneself and one’s own selfish desires. This concern with transcending one’s selfish perspective into a more rational, objective one that can be shared with other agents is reflected in each different formulation of the Categorical Imperative. Conversely, violations of that imperative are usually described in selfish terms: we put ourselves ahead of others, make an exception in favour of ourselves, or use others merely as a means for our own ends. It is no exaggeration to say that for Kant, our tendency to selfishness is the root of all evil.[[3]](#footnote-3)

 The name Kant gives to that tendency is *Selbstliebe*: self-love. He connects self-love with the pursuit of one’s own happiness.[[4]](#footnote-4) It prominently features in his work as one of the two incentives of practical reason and as a counterpart to respect for the moral law. It is thus responsible for our frequent opposition to, or inability to live up to, the demands of morality. Given the stark contrast Kant repeatedly draws between moral action and action aiming at our own happiness, it is unsurprising that many of Kant’s readers have thought that self-love *just is* selfishness and is therefore necessarily bad.[[5]](#footnote-5) On this assumption, being a morally good person requires overcoming self-love as much as possible.

 To the contrary, I will argue that Kant draws crucial distinctions among different kinds of self-love.[[6]](#footnote-6) While some are indeed selfish, and hence identified as bad, there are other forms of self-love that can be good. In particular, one form Kant identifies – as ‘self-contentment’ or, more technically, ‘rational self-love of delight – is a positive feeling that both contributes to happiness and makes the moral life more worthy of love, thereby strengthening the agent’s virtue. All this is to say that, far from being something the virtuous agent should attempt to overcome, the *right* kind of self-love is an important part of a successful moral life for Kant.

I begin with a brief discussion of the role self-love plays as an incentive in Kant’s theory of action in section 1. In section 2 I discuss Kant’s distinction between practical and pathological love, before introducing his general division between love of benevolence and love of delight[[7]](#footnote-7) in section 3. I then discuss two of the most important passages from the published work on self-love in light of this background. First, in section 4 I cover the taxonomy of self-love offered in the *Critique of Practical Reason*, most notably including self-conceit as an immoral self-love of delight. Yet since it offers no account of ‘rational’ or moral self-love of delight, this taxonomy leaves a notable gap. A rich passage from the *Religion*, which I analyse in section 5, finally provides such an account. In section 6 I conclude by examining the role this rational self-love plays in the moral life and suggest that it provides considerable philosophical benefits for Kant’s view of that life.

1. **Self-love in Kant’s action theory**

Perhaps Kant’s most familiar usage of the term ‘self-love’ is in the context of his theory of action. Here, it refers to an incentive (*Triebfeder*) of practical reason, or to the principle which incorporates that incentive as its determining ground. It is identified with the pursuit of one’s own happiness, understood as the more or less systematic attempt to satisfy one’s sensible inclinations. This story is present from the *Groundwork* onward but gets fleshed out particularly in the *Religion*. There, Kant describes how an agent’s choice to have either a good or evil fundamental maxim or *Gesinnung* is ultimately about “the incentives that he incorporates into his maxim”. Agents establish an evil *Gesinnung* when they make “the incentive of self-love and its inclinations into a condition for compliance with the moral law”[[8]](#footnote-8) rather than vice versa, as they morally ought to.

 Self-love in this action-theoretic sense thus appears primarily as a competitor to practical reason’s other incentive: the moral law.[[9]](#footnote-9) Self-love’s proper role is *just* to be subordinated to the moral law. We are not required (or able) to eliminate it entirely, and the pursuit of our happiness under moral conditions is morally permitted. It can even be indirectly morally required insofar as it helps us do our duty and avoid temptations to the contrary (KpV, AA 05: 93). That said, most of us hardly need to be encouraged to seek our own happiness. Instead, we should realise that even when we do our duty the incentive of self-love is often secretly at play: if we examine our motives carefully, “we everywhere come up against the dear self <stößt man allenthalben auf das liebe Selbst>” (GMS, AA 04: 407). This reveals our commitment to morality to still be merely conditional. Knowing that self-love has this tendency to sneak up on us and strengthen its grip when we are not looking, it seems we ought primarily to stay vigilant and suppress self-love in favour of its morally superior competitor whenever we can.

 It is not hard to see why on this story, any suggestion of a positive role for self-love in the moral life looks suspicious. While we can quarrel about its details, I think the story is generally true as a description of Kant’s action theory. My claim, however, is that there are *feelings* of self-love which are related to, but not identical with, the *incentive* of self-love. These feelings have a place in even the best moral life, in which the principle of self-love is fully subordinated to morality. In the next section I will show how this distinction is systematically grounded in Kant’s general theory of love.

1. **Practical love and actual (pathological) love**

Kant’s most frequently noticed distinction in love is that between ‘practical’ love, which is a matter of principles, and ‘pathological’ love, which is one of feeling. When he sees fit to draw it, his focus is typically on practical love. For instance, he repeatedly points out that the Biblical love command should be understood as commanding practical love, since we cannot be commanded to have a feeling (GMS, AA 04: 399; KpV, AA 05: 83). Such passages could leave readers with the impression that Kant is dismissive of pathological in favour of practical love, holding the former to be of no moral importance. This impression is strengthened by the strongly negative connotation the word ‘pathological’ has in contemporary usage. It suggests the caricature of feeling as a kind of sickness, with only cold hard duty as its cure. Kant in fact uses the term ‘pathological’ neutrally, simply to indicate ‘involving feeling’. Still, it is also important to note that this kind of love is ‘pathological’ in the sense of being something we are *affected* by passively,[[10]](#footnote-10) rather than being under our direct control. We can at best exercise *indirect* control over it by bringing about circumstances in which we will (likely) be affected by the feeling. I will argue below that making a commitment to morality is one important way in which we do so.

Yet we tend to think of love as *essentially* involving feeling, and Kant himself sometimes uses the word ‘love’ simpliciter in a way that suggests he too thinks of feeling as the primary or ‘actual’ variant of love.[[11]](#footnote-11) Accordingly, in what follows I avoid using the word ‘pathological’ and instead refer to the feeling of love, actual love, or love *simpliciter*.[[12]](#footnote-12)

So how does this feeling of love relate to practical love? As noted, both can motivate action, but the motivating ground differs. What Kant often says about the feeling of love is that what we do from that feeling, we do *gladly* (KpV, AA 05: 83; EaD, AA 08: 338) – that is, from inclination, the satisfaction of which yields pleasure. When we act from practical love, by contrast, we perform (at least roughly) the same actions but from principle instead. In other words, practical love means adopting maxims to do what a ‘pathologically’ loving agent would do from feeling alone. Typically, in the contexts in which Kant invokes practical love, he takes for granted that this principle is a moral principle so that the agent can be characterised as acting ‘from duty’. Importantly however, we can act from bad principles,[[13]](#footnote-13) and it should be possible for an act to be performed from practical love but not fromduty. Kant also thinks practical love serves to cultivate the feeling of love, though he never fully spells out the underlying mechanism.[[14]](#footnote-14)

These points about practical love let us make better sense of the distinction drawn in section 1, between self-love as an incentive for action and as a feeling. We can now understand the former as *practical self-love*. When we act on self-love in this sense, we take the natural loving attitude we have towards ourselves and elevate it into a *principle* (the “principle of one’s own happiness”). It is this principle which, since it can compete directly with the moral principle, has to be subordinated to the latter. Kant never has reason to make explicit that acting on the principle of one’s own happiness is a form of practical love. After all, he only invokes the concept of practical love to explain how acting in a loving way can be a duty even in the absence of corresponding inclinations. But the inclinations that underlie the principle of self-love are rarely, if ever, absent. We do not have a duty to form the corresponding principle, since we will do so regardless. In fact, as we saw, that principle frequently speaks *against* duty. Kant thus never has occasion to recommend a maxim of practical self-love[[15]](#footnote-15) as he recommends dutiful practical love for others.

If my suggestion that the principle of happiness is practical self-love is correct, it demonstrates the systematicity of Kant’s taxonomy of love and how having this taxonomy clearly in view can make sense of his different ways of talking about it. However, the analysis of the feeling of self-love below does not depend on it. Kant clearly has a concept of self-love as a feeling; he can have an analysis of that feeling even if he neglected to clearly identify its practical counterpart. From here on, the feeling of self-love will be our focus; we can set practical love aside except where it helps us understand that feeling.

1. **Love of benevolence, love of delight**

A second distinction in Kant’s theory of love has received considerably less attention. This is the division[[16]](#footnote-16) between *Liebe des Wohlwollens* (love of benevolence) and *Liebe des Wohlgefallens* (love of delight).[[17]](#footnote-17) It appears as early as the 1784 Collins lecture notes:

“All love is either love of benevolence, or love of delight. Love of benevolence consists in the wish and inclination to promote the happiness of others. The love of delight is the pleasure we take in showing approval of another’s perfections.”[[18]](#footnote-18)

While this early statement of the general division appears to be restricted to love of others, [[19]](#footnote-19) Kant’s considered view is that it applies equally to self-love.[[20]](#footnote-20)

This division allows Kant to capture different kinds of love. Love of benevolence is the desire for the well-being of the beloved. It motivates us to act for their benefit. When they flourish, we feel pleasure; when they suffer, we feel pain. To feel love of benevolence for someone is *to make their ends my own*. In fact, love of benevolence has a practical variety which is explicitly characterised in these terms: “The duty of love for one’s neighbour can (…) be expressed as a duty to make others’ *ends* my own”[[21]](#footnote-21). Just prior to this passage, Kant has characterised the duties of love as duties to adopt “a maxim of benevolence <Maxime des Wohlwollens>” (MS, AA 06: 449). However, there is still a difference between adopting the ends of others for oneself as grounds for action, which is the practical love that duty enjoins, and actually *feeling* emotionally invested in these ends as if they were my own. Kant himself makes explicit that the former will eventually lead to the latter: “If someone practices [beneficence] often and succeeds in realising his beneficent intention, he eventually comes actually to love the person he has helped.”[[22]](#footnote-22) Only the latter feeling is actuallove of benevolence.[[23]](#footnote-23)

If love of benevolence means taking the beloved’s ends as my own, then we can see why self-love of benevolence is natural and “happens unavoidably”. In some sense, it is trivial that agents take their ends as their own.[[24]](#footnote-24) It does not escape Kant that the Biblical love command presupposes this self-benevolence and asks us to love our neighbour ‘*as yourself*’.[[25]](#footnote-25) Self-love of benevolence becomes morally permitted “on the condition of your being benevolent to every other as well”[[26]](#footnote-26). You should love yourself because, *like others*, you are a moral agent worthy of having their ends taken seriously.

Love of delight captures a different aspect of love: the feeling of pleasure in approving of the beloved. The delight we have in our approval of them can come apart from wanting to act for their benefit. It is rather a response to the perceived perfection or worth of its object. For example, imagine your romantic partner comes home from work and starts telling you about their day working at the hospital. As they talk you find yourself struck again not just by how attractive they are, but how intelligent, kind, and driven. There is unique pleasure in this thought, and while you *do* want the best for them and are invested in their happiness, your current feeling of pleasure just isn’t about that. You are just appreciating their qualities. While there may be a healthy helping of “I can’t believe *I’m* with them” mixed in, the feeling also is not just pride that someone so great chose youas their partner. You remember feeling this way when you got to know them, before you ever dated and their qualities reflected on you in any way. What you feel is love of delight. It may motivate you to act in some scenarios to preserve your partner’s qualities or your appreciation of them but often calls for no action but to continue appreciating them.

However, since the two feelings of love are distinct, they can come apart. Kant makes this point explicitly about self-love: “The love of delight (*complacentia*) and that of benevolence to oneself (*benevolentia*), are not always united with one another, though they should be.”[[27]](#footnote-27) His phrasing suggests that a conflict between the two can be a problem, and it is easy to imagine scenarios where this would be so. For instance, imagine your perception of your partner’s worth is *too* wrapped up in their successful but overdemanding career at the hospital. If they are unhappy in that career and want to move to a less prestigious but more fulfilling job, there is now an obvious tension between your love of benevolence and love of delight. Love of benevolence enjoins you to support them doing what they really want to do; love of delight warns you that this will decrease their worth in your eyes, and thus the pleasure you feel in appreciating them.

Imagining the two feelings coming apart in other ways may provide us with intriguing tools to describe various dysfunctional love relationships. The more relevant point here, however, is that love of delight can be apt or not, depending on whether the lover has a correct perception of the worth of the beloved. If their estimation of the beloved’s worth rests on the wrong grounds, we should expect trouble to result. It will turn out below that this is exactly what happens in the case of self-love of delight. Yet this should not blind us to the real possibility of a benign case, in which the agent’s self-love is apt and based on a well-founded appraisal of their own worth. With this general framework fully in place, we can turn to Kant’s published discussions of self-love.

1. **Self-love in the second *Critique***

**4.1 Kant’s first roadmap of self-love**

A rich passage from *KpV* lays out crucial aspects of the relationship between self-love and morality. I will walk through it below:

“**[1]** All the inclinations together (which can be brought into a tolerable system and the satisfaction of which is then called one’s own happiness) constitute *self-absorption* <Selbstsucht> (*solipsismus*). **[2]** This is either the self-absorption of *love for oneself* <*Selbstliebe*>, a predominant *benevolence* toward oneself (*Philautia*), or that of *delight* with *oneself* (*Arrogantia*). The former is called, in particular, *self-love* <*Eigenliebe*>; the latter, *self-conceit*. <*Eigendünkel*>. **[3]** Pure practical reason merely infringes uponself-love <*Eigenliebe*>, inasmuch as it only restricts it (…) to the condition of agreement with this law, and then it is called *rational self-love* <*vernünftige Selbstliebe*>. But it strikes downself-conceit altogether, since all claims to esteem for oneself <Selbstschätzung> that precede accord with the moral law are null and quite unwarranted, because certainty of an orientation <Gesinnung> in accord with this law is the first condition of any worth of a person”.[[28]](#footnote-28)

Kant’s familiar starting point [1] is that pursuing one’s inclinations systematically, under the guise of happiness, constitutes a kind of selfishness here called ‘Selbstsucht’. While difficult to translate literally, this word certainly has a negative connotation – ‘-Sucht’ connotes addiction, mania, or greed.[[29]](#footnote-29) ‘Self-absorption’ captures how the agent is not merely centrally concerned with their happiness, but is so at the expense of other concerns.

In [2] Kant invokes the division into (love of) benevolence and delight we saw above to establish that there are two kinds of *Selbstsucht.*[[30]](#footnote-30)Self-love of benevolence is a ‘predominant’ benevolence to oneself, reasonably glossed as a tendency mostly to pursue one’s own ends instead of those of others. Self-love of delight is identified with self-conceit, of which no gloss is immediately given. Kant also appears to introduce a new distinction between *Selbstliebe* and *Eigenliebe*,but this turns out to be a red herring. The text does not bear out a consistent distinction,[[31]](#footnote-31) and *Eigenliebe* drops out of Kant’s vocabulary altogether after *KpV*.

Finally, in [3] Kant describes how pure practical reason’s consciousness of the moral law[[32]](#footnote-32) affects self-love. Because he has distinguished two varieties of such love, pure practical reason will have two different effects. Love of benevolence can remain. It is good to take an interest in oneself and promote one’s well-being – just not unconditionallygood. The condition of its goodness is agreement with the law. Self-love of benevolence is thus *restricted* and becomes conditioned by the moral law. It is then characterised as *rational*.

Self-conceit, however, cannot coexist with pure practical reason and is instead ‘struck down altogether.’ There is no mention here of a rational form of it. Kant’s explanation is that self-conceit is a kind of claim to ‘esteem’, where esteem is a response to a person’s perceived worth. But real worth is conditional on a person’s moral *Gesinnung*. The claims of self-conceit take no account of this, which is why they are *all* unwarranted. They are only obstacles to be removed for morality to properly take root.

Taking stock, agents begin from *Selbstsucht* – undue self-absorption, which can take the form both of over-prioritising one’s well-being and of claiming undeserved esteem for oneself. Both are kinds of self-love. However, pure practical reason demands a move away from *Selbstsucht*, restricting self-benevolence and entirely eliminating self-conceit. The passage thus gives the impression that where the agent previously had access to two kinds of self-love, morality leaves only one: ‘rational’ self-love of benevolence. Self-love of delight has appeared only in its selfish, immoral form as self-conceit: the passage leaves open whether self-conceit has a rational counterpart. Figure 1 represents what we have so far:

*Figure 1: Kant’s incomplete quadrant of self-love*

|  |  |  |
| --- | --- | --- |
|  | **Love of benevolence** | **Love of delight**  |
| **Immoral [Selfish/*Selbstsuchtig*]** | Unrestricted self-benevolence | Self-conceit |
| **Moral[rational]** | Conditioned self-benevolence | **??** |

In the next section I examine self-conceit more closely. Understanding what it is, and what is wrong with it, will bring us in a position to appreciate what the rational counterpart that will eventually complete this picture should look like.

**4.2 Self-conceit as self-love of delight**

Self-conceit has enjoyed considerable scholarly attention, though it has not usually been identified as a variety of love.[[33]](#footnote-33) Kant defines it shortly after the above-quoted *KpV* passage as “self-love making itself lawgiving and the unconditional practical principle”.[[34]](#footnote-34) This definition makes no mention of feeling or *Wohlgefallen*. It suggests instead a thin reading resonant with Kant’s view of the evil *Gesinnung* in the *Religion*: an agent is self-conceited if they prioritise self-love (of benevolence) over the moral law in their most fundamental principle.[[35]](#footnote-35) However, recall that Kant also characterised self-conceit as making (false) claims to esteem. To take oneself as sufficiently important to let one’s inclinations take priority over moral principles requires a kind of esteem for oneself that is not captured in the concept of mere benevolence to oneself.

Kant remarks in *MS* how “a lover’s wish to find only good qualities in his beloved blinds him to her obvious faults.”[[36]](#footnote-36) Love is accompanied by the desire to see the beloved object as wholly good. This makes the lover prone to illusion about the beloved’s qualities. If self-conceit is an expression of love, we should expect it to involve a similar illusion. Kant makes this point explicitly in connection with self-delight in the Vigilantius lecture notes:

“*The love of delight towards oneself*. This too is *philautia*, if it is exclusively entertained towards oneself, but also becomes unreasonable. For anyone who has liking only for himself puts himself in danger of being incapable of examining or amending his faults, and of imputing to himself, in relation to others, a certain moral worth, which he first has to acquire.”[[37]](#footnote-37)

Moran plausibly argues that a self-conceited agent does not simply reject morality. Rather, self-conceit moves agents to “construct a fiction” of being morally impartial and to use rationalisation to underpin an unwarranted confidence in the correctness of their practical judgments (Moran 2014, 422). Self-conceited agents do not see themselves as evil. Instead, they view themselves as especially good. This self-conception is fragile because it rests on a delusion that requires subconscious effort to maintain – in the terminology of section 3, it fails to be apt because it is grounded in an incorrect perception of one’s own worth.[[38]](#footnote-38) To sustain the delusion, conceited agents must be arrogant, i.e., take little or no account of the judgments of others. They interpret evidence of their worth to their own liking and are eager to take moral credit (associated with delight) wherever they can (427).[[39]](#footnote-39)

Note that in *KpV*’s Dialectic, Kant already introduces the possibility of a different, moral kind of self-delight. There he states that the “consciousness of a determination of the will directly by the law <Bewußtsein der Bestimmung des Willens unmittelbar durchs Gesetz>” is a ground of *Wohlgefallen* with oneself (KpV, AA 05:116) and elaborates that this feeling can be called “*contentment with oneself*, which in its strict meaning always designates only a negative delight in one’s existence, in which one is conscious of needing nothing.”[[40]](#footnote-40) This already suggests that self-conceit is *not* the only possible variety of self-love of delight. Unfortunately, it is not clear whether Kant here sees *Selbstzufriedenheit* as a type of love, and he gives little indication of its systematic position in the previous discussion of self-love. The delight is characterised as “only negative”; it is merely an “analogue of happiness” and “can be called intellectual contentment.”[[41]](#footnote-41) This makes it difficult to characterise as a kind of love that can really (positively) replace self-conceit.[[42]](#footnote-42)

1. **Self-love of delight in the *Religion***

In the *Religion* Kant still calls self-love “the source of all evil <die Quelle alles Bösen>” (RGV, AA 06: 45). Ironically, however, it is in a footnote[[43]](#footnote-43) to that very sentence that he provides his most nuanced positive discussion of it:

“**[1]** Like love in general, **self-love**too can be divided into love of **benevolence** and of **delight** (benevolentiae et complacentiae), and both must, as is self-evident, be rational. (…) **[2]** A rational <vernünftige> love of delight for oneself can either be understood as a delight in the aforementioned maxims which have as their end the gratification of natural inclinations; and then, it is one and the same <einerlei> with love of benevolence for oneself. (…) **[3]** But the maxim of self-love of an unconditional (not dependent on profit or loss as the consequences of action) *delight* in oneself would be the inner principle of a contentment only possible for us under the condition of the subordination of our maxims to the moral law. **[4]** No human being who is not indifferent to the moral law and who is conscious of such maxims as do not agree with the moral law inside of him, can have a delight in himself, indeed even be without a bitter dissatisfaction <Mißfallen> in himself.One could call this **rational love** <Vernunftliebe> of oneself, which prevents all admixture of other causes of contentment from the consequences of one’s actions (under the guise of a happiness which is to be obtained thereby) with the incentives of volition.”[[44]](#footnote-44)

Kant makes clear in [1] that the general division applies to self-love, then immediately adds to his previous position. He now holds that *both* types of self-love must be rational, implying a rational variety of self-love of delight not mentioned in the *KpV* passage. After another discussion of self-love of benevolence, omitted here,[[45]](#footnote-45) Kant turns to self-love of delight in [2]. He first introduces an instrumentally rational sense of self-love of delight. He then says it is “einerlei” with self-love of benevolence.[[46]](#footnote-46) It soon becomes clear, however, that this is not the sense of ‘rational’ he is primarily interested in.

Instead, in [3] we are finally introduced to a new, *unconditional* kind of self-delight. Agents who tie their delight in themselves to the results of their actions leave it partly up to factors outside their own control. What isalways up to them, however, is the choice to make the moral law the supreme condition of their maxims. This choice can ground an unrestricted delight. Kant then calls this delight a kind of *love*: *Vernunftliebe*. This love appears to be ‘rational’[[47]](#footnote-47) not in the prudential sense, but in the sense of being conditioned by the moral law – the same sense in which conditional self-love of benevolence was in *KpV*. Agents with a good *Gesinnung* feel an unconditional love of delight for themselves, which is again characterised as contentment.

Furthermore, with [4] we get a more complete picture of the affective process of the moral law ‘striking down’ self-conceit. First, consciousness of the moral law generates an uncomfortable ‘dissatisfaction’ with ourselves which makes it harder to sustain our unjustified self-delight. This helps strike it down. If we successfully undergo moral transformation and achieve *Vernunftliebe*, however, that love also performs a morally motivating role. It prevents us from acting for the sake of this false self-love of delight. It thus ‘keeps down’ the latter, but *not* through dissatisfaction – rather, by offering us something better, dampening our motivation to pursue what is worse.

The footnote goes on to reaffirm that this love cannot be the objective condition of our commitment to morality and is conditional on our *worthiness* to be happy; yet this is not to say that, once we are so worthy, our subjective wish for happiness cannot play a motivating role. Note a crucial slide here. Kant no longer speaks of self-contentment as providing a mere “analogue” of happiness. Self-contentment works because it *does* satisfy our wish for happiness and provides a happiness of which the morally good agent is worthy. There is no indication that the delight of self-contentment is “negative”, and while it is “rational” this does not imply that it is merely intellectual. What seems to explain this change is that self-contentment has now been properly brought into the fold as a kind of *love*, thus completing the picture provided in section 4:

*Figure 2: Kant’s quadrant of self-love, completed.*

|  |  |  |
| --- | --- | --- |
|  | **Love of benevolence** | **Love of delight**  |
| **Immoral [Selfish/*Selbstsuchtig*]** | Unrestricted self-benevolence | Self-conceit |
| **Moral[rational]** | Conditioned self-benevolence | *Self-contentment* |

1. **The role of rational self-love**

We now see that, despite appearances, Kant really is committed to the existence of rational self-love of delight, and therefore to a kind of self-love available only to good agents. Admittedly he does not say much to explain why, but we can see the philosophical benefits of this view by developing it in the following way.

 Firstly, there is a distinction between the choice to commit to the moral law as one’s primary incentive, and *virtue*. Kant enriches his description of the former significantly in the *Religion*, where he first begins to characterise it as a “change of heart”[[48]](#footnote-48) (*RGV*: 6:47), but never abandons the central claim of his moral theory that this choice must be made autonomously, therefore not based on some sensible feeling or hoped-for result. It would be problematic for Kant suddenly to attribute a motivating role to self-love of delight in the choice to be moral, inviting the charge of heteronomy. There is no reason to believe that he does.

 However, Kant increasingly emphasises in his later works[[49]](#footnote-49) that even agents who choose to be moral are not thereby suddenly rid of contrary inclinations. They need the strength to reliably act in line with this choice, and to do their duty even in the face of those inclinations. This strength is virtue (cf. MS, AA 06: 384, 394, 405). Virtue must be able to directly compete with, and overcome, inclinations, and Kant never pretends that this is all a matter of intellect alone. Virtue has an affective dimension, and the more powerful the agent’s feelings in favour of morality, the better they will be able to resist feelings that pull in the other direction.[[50]](#footnote-50) Kant accordingly claims that the “aesthetic constitution, as it were the *temperament* *of virtue*”, ought not be “downcast and fearful” but “joyful”. The fearful attitude “cannot occur without a hidden *hatred* of the law”, while “a heart joyful in the observance of its duty” is evidence that the agent has acquired “a *love* for the good”.[[51]](#footnote-51) Kant here reveals a clear concern that the moral law be not merely an authoritative command worth obeying but also worth *loving*. All else being equal, the agent who loves the law will be more virtuous than the one who follows it in quiet resentment. After all, the former’s feelings will be an aid, rather than a hindrance, to the force of their commitment to morality.

A similar point is made in *The End of All Things*, which characterises “true love” as an “indispensable complement to the imperfection of human nature”, “for what one does not do gladly, they do so paltrily that [respect] cannot be counted on for much as an incentive without being joined by [love].”[[52]](#footnote-52) Kant is happy for love to be a subjective ground for doing one’s duty, even as “respect is without a doubt primary <Die Achtung ist ohne Zweifel das Erste>.” So, while respect remains the only incentive that can lead the agent to choose to commit to morality, it is perfectly reasonable for the agent to require the additional motivation of love in order to strengthen their virtue and persist in their commitment to the moral life.[[53]](#footnote-53)

It is hard to imagine how the moral life could be worthy of love, however, if it would not even let us love ourselves. As our tendency to self-conceit illustrates, human beings have a deep need to do so. There is no reason to think this need would evaporate just because we become better people. A ‘merely negative’ self-contentment is just an absence of feeling and cannot fulfil that need. Rational self-love of delight can. This love is itself a positive feeling and source of pleasure. Moreover, it is essentially connected to the moral law – we find that we love ourselves precisely *insofar* as we are living up to our moral vocation.[[54]](#footnote-54) In connecting morality with happiness, self-love of delight helps us do our duty gladly – therefore making morality itself more worthy of love. Rather than placing us at odds with morality and tempting us to evil, self-love now appears as an indispensable complement to a moral life worth living for agents like ourselves – both capable of transcending selfishness in the service of duty *and* still in need of happiness and love.

Note how this brings Kant significantly more in touch with how many of us think of self-love today. We are often encouraged to love ourselves. We are told we deserve to, or that unless we learn to love ourselves, we cannot expect to be any good at loving and being loved by others. These encouragements happen largely because many of us still struggle to love ourselves. Loving oneself in this sense appears to be a significant moral achievement, not a state we unavoidably find ourselves in. Conversely, it is common to be selfish without *really* loving oneself.

Kant as I have sketched him is in a better position to capture these insights than he appeared on the standard picture. We *should* indeed love ourselves, rationally. Doing so will help us not just relate better to ourselves, but also to others – at minimum, in the sense that it will strengthen our ability to perform our duties towards them, including the duties of love. And loving oneself rightly is *hard*: it requires overcoming deep-seated, selfish kinds of love and changing one’s heart to genuinely commit to morality. Kant’s view has the added benefit of emphasising a move from dysfunctional to ‘true’ or healthy varieties of love, while insisting that both really count as love.[[55]](#footnote-55)

1. **Conclusion**

To conclude, I have argued against a common picture according to which Kant sees self-love as nothing but selfishness, and therefore bad altogether. Instead, I have shown that Kant develops a sophisticated vocabulary which allows him to distinguish bad (self-absorbed) from good (rational) varieties of self-love. While morality demands we overcome the former, the latter play an important role in moral life.

First, Kant distinguishes the feeling of self-love from self-love as practical principle and antagonist of the moral law. Then, the division between love of benevolence and love of delight lets him distinguish different feelings, which interact differently with morality. Self-love of benevolence is our tendency to be emotionally invested in our own ends. This feeling needs only to be restricted, from a self-absorbed to a rational, morally conditioned benevolence. Self-love of delight first appears in its guise of self-conceit, which is an esteem for ourselves based on an illusory conception of our worth. Because it locates this worth *not* in its real source – morality – but outside it, self-conceit forms an obstacle to morality and needs to be uprooted altogether.

Yet, once Kant fully appreciates the need for morality to be supplemented by love, he also articulates an ideal of rational self-love of delight. This is a self-love that is only available to morally good agents based on an accurate perception of their own worth. Loving themselves in this way allows them to overcome the need for self-conceit by finding self-contentment, and powerfully aids virtue by strengthening inclinations that align with morality. Ultimately, it helps us come to love the moral law itself, which Kant increasingly realises is necessary for the moral life to be worth living for sensibly affected rational agents. Thus, Kantian agents are not simply torn between the moral law and self-love, forced to sacrifice one to live by the other. Good agents love themselves. While for self-absorbed agents, the gulf between self-love and the demands of morality appears unbridgeable, agents who truly commit to morality find that the moral life brings the two into unexpected harmony.

WORKS CITED

Arroyo, Christopher. (2016) “Kant on the emotion of love.” *European Journal of Philosophy*

24(3), pp. 580-606.

Baxley, Anne Margaret. (2010) *Kant’s Theory of Virtue*. Cambridge: Cambridge University Press.

Borges, Maria. (2012) “A typology of love in Kant’s philosophy.” In Frederick Rauscher and

Daniel Omar Perez (eds.), *Kant in Brazil* (Rochester: University of Rochester Press), pp. 271-282.

(2019) *Emotion, Reason, and Action in Kant*. London: Bloomsbury Academic.

Cohen, Alix. (2020) “A Kantian Account of Emotions as Feelings”. *Mind* 129 (514): pp. 429-60.

Deimling, Wiebke. (2018) “Two Different Kinds of Value? Kant on Feeling and Moral Cognition”. In

Kelly Sorensen & Diane Williamson (eds.), *Kant and the Faculty of Feeling*, pp. 25–40. Cambridge: Cambridge University Press.

Deligiorgi, Katerina. (2012) *The Scope of Autonomy: Kant and the Morality of Freedom*. Oxford: Oxford University Press.

DeWitt, Janelle. (2018) “Feeling and Inclination: Rationalizing the Animal Within”. In Sorensen &

Williamson (eds.), *Kant and the Faculty of Feeling*, pp. 67-87.

Elizondo, E. Sonny. (2014) “More than a feeling.” *Canadian Journal of Philosophy* 44(3-4): pp. 425-442.

 (2016) “Morality is its own reward.” *Kantian Review* 21(3): pp. 343-365.

Frankfurt, Harry. (2001) “The dear self”. *Philosopher’s Imprint* 1(0), pp. 1-14.

(2004) *The Reasons of Love*. Princeton: Princeton University Press.

Frierson, Patrick R. (2014) *Kant’s Empirical Psychology*. 1st ed. Cambridge: Cambridge University Press.

Grenberg, Jeanine. (2014) “All you need is love?” In Alix Cohen (ed.), *Kant on Emotion and Value*, pp.

210-223. New York: Palgrave Macmillan.

(2015) “Love.” In Lara Denis and Oliver Sensen(eds.), *Kant’s Lectures on Ethics: A Critical Guide*, pp. 239-255. Cambridge: Cambridge University Press.

(2022) *Kant’s Deontological Eudaemonism*. Oxford: Oxford University Press.

Herman, Barbara. (1993) ‘Integrity and Impartiality’. In *The Practice of Moral Judgment*, pp. 23-44. Cambridge, Mass: Harvard University Press.

(2007) ‘Making Room for Character’. In *Moral Literacy*, pp. 1-28. Cambridge, Mass: Harvard University Press.

Hutcheson, Francis. (2004) “An inquiry concerning moral good and evil.” In Wolfgang

Leihold (ed.), *An Inquiry Into The Original of Our Ideas of Beauty and Virtue in Two Treatises*, pp. 83-198. Indianapolis: Liberty Fund.

Kant, Immanuel. (1996a) *The End of All Things*. Translated by Allen W. Wood. In Allen W.

Wood and George Di Giovanni (eds.), *Religion and Rational Theology*. Cambridge: Cambridge University Press.

(1996b) *Religion within the Boundaries of Mere Reason*. Translated by George Di Giovanni. In *Religion and Rational Theology*.

(1996c) *The Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press.

(1997) *Lectures on Ethics*. Translated by Peter Heath. Edited by Peter Heath and Jerome Schneewind. Cambridge: Cambridge University Press.

(1999) *Correspondence*. Translated and edited by Arnulf Zweig. Cambridge: Cambridge University Press.

(2000) *Critique of the Power of Judgment*. Translated by Paul Guyer and Eric Matthews. Edited by Paul Guyer. Cambridge: Cambridge University Press.

(2007a) *Anthropology from a Pragmatic Point of View*. Translated by Robert B. Louden. In Günther Zöller and Robert B. Louden (eds.), *Anthropology, History, and Education*. Cambridge: Cambridge University Press.

(2011) *Groundwork to the Metaphysics of Morals*. Translated and edited by Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press.

(2015) *Critique of Practical Reason*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press.

McCarty, Richard. (2009) *Kant’s Theory of Action*. Oxford: Oxford University Press.

Moran, Kate. (2014) “Delusions of virtue: Kant on self-conceit”. *Kantian Review* 19(3), pp. 419-447.

Muchnik, Pablo. (2014) “The heart as locus of moral struggle in the *Religion*.” In Cohen (ed.), *Kant on*

*Emotion and Value*, pp. 224-244.

Palmquist, Stephen R. 2016. *Comprehensive Commentary on Kant’s Religion Within the Bounds of Bare Reason*. Oxford: Wiley Blackwell.

Ricken, Friedo. (2001) “Kant über Selbstliebe: ‘Anlage zum Guten’ oder ‘Quelle alles Bösen’?”

*Philosophisch Jahrbuch* 108(2), pp. 245-258.

Rinne, Pärttyli. (2018) *Kant on Love*. Berlin: De Gruyter.

Rousseau, Jean-Jacques. (1973) *The Social Contract and the Discourses*. Translated by G.D.H. Cole. Edited by

J.H. Brumfitt and John C. Hall. New York: Alfred A. Knopf.

Russell, Francey. (2020) “Kantian self-conceit and the two guises of authority”. *Canadian Journal of*

*Philosophy* 50(2), pp. 68-283.

Seymour Fahmy, Melissa. (2010) “Kantian practical love.” *Pacific Philosophical Quarterly* 91, pp. 313-331.

Sherman, Nancy. (1997) *Making a Necessity of Virtue: Aristotle and Kant on Virtue*. Cambridge: Cambridge University Press.

Smith, Catherine M.M. (2017) “Must self-conceit be self-conceited?” In *Self-Respecting Animals: Three Papers*

*on Kant’s View of Human Nature and Morality*. Cornell University, PhD dissertation, pp. 9-44.

Sorensen, Kelly D. (2002) “Kant’s taxonomy of the emotions”. *Kantian Review* 6, pp. 109-128.

Ware, Owen. (2020) “Self-love and self-conceit.” Unpublished manuscript. Last modified

April 15, 2020. <https://philpapers.org/rec/WARSAS-5>.

Wood, Allen W. (1999) *Kant’s Ethical Thought*. Cambridge: Cambridge University Press.

(2020) *Kant and Religion*. Cambridge: Cambridge University Press.

1. This paper has benefitted from generous feedback and discussion by many. I would like to single out Monique Wonderly, who first encouraged me to write it, and Eric Watkins, who tirelessly provided feedback on its different iterations. I am deeply grateful to both of them. I also want to thank Lucy Allais, Rosalind Chaplin, Juan Carlos Gonzalez, Jeanine Grenberg, Monte Johnson, James Messina, Dana Nelkin, Konstantin Pollok, Sam Rickless, Pärttyli Rinne, Donald Rutherford, Joseph Stratmann, Clinton Tolley, and the audience at the Eastern NAKS Study Group 2019 hosted by Kate Moran at Brandeis University. [↑](#footnote-ref-1)
2. I have consulted the Cambridge Kant translations, but the translations here are my own. [↑](#footnote-ref-2)
3. Cf. RGV, AA 06: 45. This does not imply that every evil action is itself a direct expression of selfishness. [↑](#footnote-ref-3)
4. Theorem II of KpV treats “the principle of self-love or one’s own happiness <das allgemeine Princip der Selbstliebe oder eigenen Glückseligkeit>” (KpV, AA 05: 22) as equivalent descriptions. [↑](#footnote-ref-4)
5. Harry Frankfurt argues that Kant gives self-love a bad name by confusing it with “mere self-indulgence” in Frankfurt 2001, 3. See also Frankfurt 200, 68. [↑](#footnote-ref-5)
6. Few attempts have been made to systematically examine the, admittedly scattered, distinctions Kant draws between different kinds of love. The notable exception here is Rinne 2018, whose work has been helpful in this regard. Borges also provides a “typology” of love, but hers is focused primarily on the discussion of “duties of love” in *MS*, not on the feeling of love for oneself (Borges 2012). Finally, Arroyo also systematically analyses many of Kant’s distinctions, but his view that love is “an expression of our unsocial sociability” (Arroyo 2016, 593) excludes self-love entirely and treats it as synonymous with self-interest (584), rather than appreciating its continuity with love of others.
Applying these distinctions to self-love, interpreters commonly go wrong by overemphasising Kant’s distinction between practical and pathological love (cf. Borges, 2012; Arroyo, 2016). In Wood’s case, it is the distinction between *Eigenliebe* and *Selbstliebe* (Wood 2020). Ricken rightly focuses on the distinction between *Liebe des Wohlwollens* and *Liebe des Wohlgefallens* but mischaracterises that distinction by relying excessively on the *Religion’s* characterisation of rational self-love; relatedly, he distinguishes rational self-love insufficiently from respect (Ricken 2001, 253). [↑](#footnote-ref-6)
7. *Wohlgefallen* and its cognates in Kant are subject to a wide, often inconsistent, variety of English translations, of which ‘satisfaction’ is the most plausible alternative. However, I have been persuaded by Jeanine Grenberg to retain ‘delight’ here, since ‘satisfaction’ misleadingly implies a ‘desire-satisfaction’ model of pleasure and happiness. As I understand the love of delight, it is a kind of pleasure that does notconform to such a model and does not always need to satisfy a specific pre-existing desire of the agent. I encourage the reader to think of *Wohlgefallen* as a technical term. I thank Grenberg for insightful discussion of this issue. [↑](#footnote-ref-7)
8. “der Triebfedern, die er in seine Maxime aufnimmt (…) er die Triebfeder der Selbstliebe und ihre Neigungen zur Bedingung der Befolgung des moralischen Gesetzes macht” RGV, AA 06: 36. [↑](#footnote-ref-8)
9. Kant variously describes both the moral law itself and respect for that law as the moral incentive. For my purposes, nothing hangs on which description is more accurate. See (McCarty 2009, 33; cf. 40, 46), who argues that Kant’s notion of ‘incentive’ is ambiguous, allowing for both descriptions to be true on different disambiguations. [↑](#footnote-ref-9)
10. Etymologically, ‘pathological’ is connected to ‘passion’, which may cause worry that pathological love is a passion (*Leidenschaft*) in Kant’s sense. If so, the caricature above would have some bite after all: since passions tend to overpower reason, Kant holds that to be subject to them is “wohl immer Krankheit des Gemüths” (Anth., AA 07: 251). But Kant does not include love in his section headers describing the different passions, and while he briefly discusses a “Leidenschaft (…) der Liebe,” the discussion is very qualified, and only “leidenschaftlich Verliebtsein>” (07: 266) is explicitly called a passion. This suggests love is not itself a passion, but rather a more complex emotion which can be accompanied by a cluster of feelings, some of which may be passions. [↑](#footnote-ref-10)
11. For instance: “*Liebe* ist eine Sache der *Empfindung*, nicht des Wollens, und ich kann nicht lieben, weil ich *will*, noch weniger aber weil ich soll (zur Liebe genöthigt werden); mithin ist eine Pflicht zu lieben ein Unding. (…) Man nennt aber oftmals ein uneigennütziges Wohlwollen gegen Menschen auch (obzwar sehr uneigentlich) *Liebe* (…) Was man aber aus Zwang thut, das geschieht nicht aus Liebe.” (MS, AA 06: 401) Cf. Arroyo 2016, 586. [↑](#footnote-ref-11)
12. Kant is notoriously unclear about the nature of feelings. I take my view to be consistent with any of the various accounts of feelings on offer: see for instance Sorensen 2002, Deimling 2018, DeWitt 2018, Borges 2019, Cohen 2020. [↑](#footnote-ref-12)
13. Kant is not always clear about this, but it is implicit in the fact that we can even *have* a “principle of self-love.” It is explicit in his discussion of character in Anth*.*, AA 07: 290. I follow Frierson in thinking of character as the motivating ground that allows us to act on the representation of principles at all, moral or otherwise (Frierson 2014, 83). Actions ‘from duty’ are then a subset of actions from character, namely those performed from the representation of a *moral* principle. The details of this account are not essential for my view, however, so long as it is granted that action from non-moral principles is somehow possible for Kant. [↑](#footnote-ref-13)
14. Cf. MS, AA 06: 402; Seymour Fahmy 2010, 324. [↑](#footnote-ref-14)
15. But see footnote 22. [↑](#footnote-ref-15)
16. Rinne calls this the “general division of love” (Rinne 2018, 6). [↑](#footnote-ref-16)
17. Love of benevolence is not loving benevolence itself; it rather means loving some object *in a benevolent way*. Idem for love of delight. I do not modify the term since the same ambiguity exists in the original German and Latin, as well as in the English writings of Francis Hutcheson which are likely Kant’s immediate source for the distinction. (“Love toward rational agents, is subdivided into love of complacence or esteem, and love of benevolence”. Hutcheson 2004, 102. See also Rinne 2018, 9n.) [↑](#footnote-ref-17)
18. “Alle Liebe ist entweder Liebe des Wohwollens oder Liebe des Wohlgefallens. Die Liebe des Wohlwollens besteht im Wunsch und in der Neigung das Glück anderer zu befördern. Die Liebe des Wohlgefallens ist das Vergnügen, welches wir haben, den Vollkommenheiten des andern Beyfall zu beweisen.” VE-Collins, AA 27: 417. [↑](#footnote-ref-18)
19. Hutcheson originally drew the distinction only at the level of love for others and seems to have opposed both love of benevolence and love of complacence *to* self-love (Hutcheson 2004, 103f). [↑](#footnote-ref-19)
20. He explicitly extends it to self-love in RGV, AA06: 45. [↑](#footnote-ref-20)
21. “Die Pflicht der Menschenliebe kann also auch so ausgedrückt werden: sie ist die Pflicht Anderer ihre Zwecke (…) zu den meinen zu machen”. MS, AA06: 450. [↑](#footnote-ref-21)
22. “Wer diese oft ausübt, und es gelingt ihm mit seiner wohltätigen Absicht, kommt endlich wohl gar dahin, den, welchen er wohl gethan hat, wirklich zu lieben.” MS, AA 06: 402. [↑](#footnote-ref-22)
23. I thus disagree with Arroyo, who insists that “Kant seems to equate pathological love, the feeling of love, with the ‘love [of delight]’” (Arroyo 2016, 586). [↑](#footnote-ref-23)
24. This should be carefully qualified. Sometimes, e.g., in severe trauma or mental distress, it seems people do lose interest in their own ends, and to take up such interest again is an entirely non-trivial task. We should at least keep open a conceptual space for people who are so detached from themselves that they feel no self-love of benevolence. We should want such people to engage in the same exercise of practical self-love Kant imagines for love of others. Beginning to take their own ends seriously out of duty may help them reestablish a benevolently loving connection with themselves. Kant does not explicitly seem to acknowledge the possibility of such cases. However, he was arguably familiar with at least one from his correspondence with Maria von Herbert (Br., AA 11: 273-4; 400-3). [↑](#footnote-ref-24)
25. Compare: “Die größte Liebe gegen andre ist die, wenn ich ihn so liebe als mich selbst, ich kann einen andern nicht mehr lieben als mich; wenn ich ihn aber so lieben will als mich, so kann ich dieses nicht anders thun, als wenn ich versichert bin, daß mich der andre eben so lieben wird als sich”. VE*-*Collins, AA 27: 423-4. [↑](#footnote-ref-25)
26. “geschieht unvermeidlich (…) unter der Bedingung, daß du auch jedem Anderen wohl willst”. MS, AA 06:451. [↑](#footnote-ref-26)
27. “Die Liebe des Wohlgefallens (*complacentia*) und Liebe des Wohlwollens gegen sich selbst (benevolentia) sind nicht immer mit einander vereint, ob sie gleich sollten.” VE*-*Vigilantius, AA 27: 675. [↑](#footnote-ref-27)
28. “Alle Neigungen zusammen (die auch wohl in ein erträgliches System gebracht werden können, und deren Befriedigung alsdann eigene Glückseligkeit heißt) machen die *Selbstsucht* (solipsismus) aus. Diese ist entweder

die der *Selbstliebe*, eines über alles gehenden *Wohlwollens* gegen sich selbst (Philautia), oder die des *Wohlgefallens* an sich selbst (Arrogantia). Jene heißt besonders *Eigenliebe*, diese *Eigendünkel*. Die reine praktische Vernunft thut der Eigenliebe blos Abbruch, indem sie solche (…) nur auf die Bedingung der Einstimmung mit diesem Gesetze einschränkt; da sie alsdann *vernünftige Selbstliebe* genannt wird. Aber den Eigendünkel schlägt sie gar nieder, indem alle Ansprüche der Selbstschätzung, die vor der Übereinstimmung mit dem sittlichen Gesetze vorhergehen, nichtig und ohne alle Befugniß find, indem eben die Gewißheit einer Gesinnung, die mit diesem Gesetze übereinstimmt, die erste Bedingung alles Werths der Person ist”. KpV, AA 05: 73. [↑](#footnote-ref-28)
29. I thank Konstantin Pollok for discussion of this term. [↑](#footnote-ref-29)
30. Admittedly, he explicitly characterises only benevolence to oneself as *Selbstliebe* – appearing to imply that the delight in oneself is not self-love. His language, however, is otherwise identical to that used in the lectures for the general division (pairing *Wohlwollen* with *philautia* and *Wohlgefallen* with *arrogantia*). It seems he uses ‘self-love’ here in a narrow sense to contrast with self-conceit, while as we will see below, his considered view regards both *Philautia* and *Arrogantia* as forms of Selbstliebe (cf. Rinne 2018, 43). [↑](#footnote-ref-30)
31. Wood attempts to make the distinction consistent by drawing a parallel with the one Rousseau makes between *amour-de-soi* [of-self; Selbst]and *amour-propre* [self-; Eigen]. The former is a natural drive to self-preservation, which “guided in man by reason (…) creates humanity and virtue.” The latter is the drive to compare oneself favourably to others; it is social in origin, “purely relative,” and the source of vice (Rousseau 1973, 74). Wood’s discussion of the predisposition to humanity brings the Rousseauian current in Kant directly to bear on the *Selbst*-/*Eigenliebe* distinction: “natural self-love (*Selbstliebe*) assumes a different, specifically rational form: *Eigenliebe*, *philautia*, which is associated with an evaluation of oneself and a propensity to place a higher value on oneself than one does on others, which is self-conceit (*Eigendünkel*, *arrogantia*)” (Wood 2020, 71).

Plausible though this parallel would be, Kant simply does not use *Eigenliebe* in this way. In the *Religion* Kant consistently uses only *Selbstliebe*. Even in the above passage, Kant does not associate *Eigenliebe* with self-conceit but directly contrasts them. If I am correct that one is love of benevolence, and the other love of delight, then this is as it should be. Moreover, Wood claims that *Eigenliebe* is the rational form of *Selbstliebe*; but in the next sentence of the passage, when Kant considers the rational form of *Eigenliebe*, he switches back to *Selbstliebe*. [↑](#footnote-ref-31)
32. This consciousness, though all agents have it, is not always effective. In the *Religion* Kant more clearly emphasises that a moral transformation (“change of heart”) is required to restore the moral law to its proper place in our hearts. Only once pure practical reason is empowered by this transformation is it able to do what it is here described as unavoidably striving towards. [↑](#footnote-ref-32)
33. Russell 2020, 272-4 gives an overview of the interpretive debate. [↑](#footnote-ref-33)
34. “wenn [Selbstliebe] sich gesetzgebend und zum unbedingten praktischen Princip macht”. KpV, AA 05: 74. [↑](#footnote-ref-34)
35. Rinne contends on this basis that although Kant distinguishes moral self-benevolence and self-delight, he identifies immoral self-delight (as self-conceit) directly with immoral self-benevolence (Rinne 2018, 40). [↑](#footnote-ref-35)
36. “wie der Wunsch eines Liebhabers lauter gute Eigenschaften an seiner Geliebten zu finden ihm ihre augenscheinliche Fehler unsichtbar macht”. MS, AA 06: 430. [↑](#footnote-ref-36)
37. “*In die Liebe des Wohlgefallens gegen sich selbst*. Auch diese ist *Philautie*, wenn sie ausschließend gegen sich selbst ausgeübt wird, wird aber auch vernunftwidrig. Denn wer sich selbst allein gefällt, setzt sich in Gefahr, daß er außer Stande ist, seine Fehler zu prüfen, und zu bessern, sowie sich im Verhältniß gegen andere einen gewissen moralischen Werth anzudichten, den er sich erst erwerben muß.” VE-Vigilantius, AA 27: 621. [↑](#footnote-ref-37)
38. Owen Ware rightly notes that when the conceited agent’s self-image is threatened by the moral law striking it down, it is likely to ‘strike back.’ Left unchecked, this can result in a further tendency to rationalise one’s way out of compliance with the moral law and thereby to further evil. Ware, however, goes too far in suggesting, paradoxically, that “this is where radical evil might arise” (Ware 2020). Radical evil, by definition, is at the root of our evil tendencies – it is the ground of our tendency to see ourselves as better than we are and cannot be its result in turn. [↑](#footnote-ref-38)
39. This description does not apply equally to all morally deficient agents. Surely not all such agents love themselves, even poorly – plenty appear to be lacking in self-esteem or to be positively self-hating. I omit full discussion of this issue here but note that self-love of delight need not always occur in full force. Especially since it can be struck down without necessarily being replaced, there is room for an (unstable) intermediary state in which agents lack *any* self-delight. We still often think there is something (perhaps excusably) *selfish* in the attitude of such agents. Whether this can still be described as self-conceited or not is interesting but not essential to my present purposes. Cf. Smith 2017 for a sensitive discussion of this problem. [↑](#footnote-ref-39)
40. “*Selbstzufriedenheit*, welches in seiner eigentlichen Bedeutung jederzeit nur ein negatives Wohlgefallen an seiner Existenz andeutet, in welchen man nichts zu bedürfen sich bewußt ist.” KpV, AA 05: 117. [↑](#footnote-ref-40)
41. “Analogon der Glückseligkeit (…) diese [Zufriedenheit] kann intellectuell heißen.” KpV, AA 05: 117-8. [↑](#footnote-ref-41)
42. Elizondo has argued for self-contentment as a purely intellectual feeling that, while it is *not* happiness, connects morality with “well-being” (Elizondo 2014, 2016). Though he does not directly address self-conceit, one might take his argument to support the view that intellectual self-contentment *is* an adequate counterweight to it. It certainly would imply that no ‘pathological’ feeling is required to make the moral life worth living. I lack space to address his argument in the detail it deserves but broadly endorse Grenberg’s qualified rejection of it. In brief, she argues that it violates the strictures of transcendental idealism because it would mean we can have a feeling directly based on freedom without being sensibly affected (Grenberg 2022, 303). Kant’s increasing interest in love in the 1790s, discussed in section 6, is further evidence that by then he believed sensibly affected rational agents need a full-blooded sensible feeling of love related to happiness proper, not a merely intellectual contentment in doing the right thing. [↑](#footnote-ref-42)
43. That it happens in a footnote does not mean it is unimportant. Kant appears to want the main text of the *Religion* to be readable on its own terms, while using its dense footnotes to provide technical detail tying it into his wider, systematic views. [↑](#footnote-ref-43)
44. “Wie Liebe überhaupt, so kann auch **Selbstliebe** in die des **Wohlwollens** und des **Wohlgefallens** (benevolentiae et complacentiae) eingetheilt werden, und beide müssen (wie sich von selbst versteht) vernünftig sein. (…) Eine vernünftige Liebe des Wohlgefallens an sich selbst kann nun entweder so verstanden werden, daß wir uns in jenen schon genannten auf Befriedigung der Naturneigung abzweckenden Maximen (so fern jener Zweck durch Befolgung derselben erreicht wird) Wohlgefallen; und da ist sie mit der Liebe des Wohlwollens gegen sich selbst einerlei; (…) Allein die Maxime der Selbstliebe des unbedingten (nicht von Gewinn oder Verlust als den Folgen der Handlung abhängenden) *Wohlgefallens* an sich selbst würde das innere Princip einer allein unter der Bedingung der Unterordnung unserer Maximen unter das moralische Gesetz uns möglichen Zufriedenheit sein. Kein Mensch, dem die Moralität nicht gleichgültig ist, kann an sich ein Wohlgefallen haben, ja gar ohne ein bitteres Mißfallen an sich selbst sein, der sich solcher Maximen bewußt ist, die mit dem moralischen Gesetze in ihm nicht übereinstimmen. Man könnte diese die Vernunftliebe seiner selbst nennen, welche alle Vermischung anderer Ursachen der Zufriedenheit aus den Folgen seiner Handlungen (unter dem Namen einer dadurch sich zu verschaffenden Glückseligkeit) mit den Triebfedern der Willkür verhindert.” RGV, AA06: 45-6f. [↑](#footnote-ref-44)
45. In the omitted passage Kant identifies rational love of self-love of benevolence in terms apparently conflicting with those in *KpV*. Here, it seem self-benevolence is rational when it is incorporated in the agent’s maxim in a way that effectively maximises their happiness. Ricken takes this as evidence that self-love of benevolence has no particular relation to morality at all (Ricken 2001, 253). Yet Kant gives no warning of having gone back on his earlier characterisation of rational self-benevolence as conditioned by the moral law. Statements to that effect also still appear in MS, AA 06: 451. More likely, Kant is using ‘rational’ in two different senses, the first moral and the second prudential. His discussion of rational self-love of delight, under [3], corroborates this. [↑](#footnote-ref-45)
46. This might be taken to support Rinne’s contention that self-conceit is identical to immoral self-benevolence (Rinne 2018, 40). “Einerlei” however can also indicate a weaker sense of identity, or just that the differences between two things are irrelevant to present purposes. [↑](#footnote-ref-46)
47. While both are best translated ‘rational love’, Kant earlier used ‘vernünftige Liebe’ perhaps to mark this contrast with ‘Vernünftliebe’. [↑](#footnote-ref-47)
48. Kant’s shift to a systematic use of the notion of ‘heart’ in the *Religion* and onward seems more significant to me than it has usually been taken to be. Kant quietly appropriates Pietist language without tipping his hand to his Pietist readers that he is using the term differently. It indicates his increased realisation that a complete moral revolution has *both* a rational and an emotional component, which goes hand in hand with the story I tell below. If morality is to work upon the heart, it makes sense for Kant to increasingly emphasise the role of love within it. However, this deserves more attention than I can give it here, and nothing in my account hangs on it. Two commentators who draw attention to Kant’s usage of ‘heart’ are Muchnik 2014 and Palmquist 2016. [↑](#footnote-ref-48)
49. Whether this reflects Kant substantially refining or changing his view in the 1790s or a mere change in emphasis between works is beyond my purview here. [↑](#footnote-ref-49)
50. I take myself here to be in broad agreement with, and contributing to, an interpretive tradition that emphasises a ‘warmer’ Kant by drawing attention to the affective and virtue-theoretic elements of his moral thought. This tradition is represented, among others, by Sherman 1997, Herman 1993 and Herman 2007, Wood 1999, Baxley 2010, Deligiorgi 2012, and Grenberg 2022. [↑](#footnote-ref-50)
51. “die ästhetische Beschaffenheit, gleichsam das Temperament der Tugend (…) ängstlich-gebeucht und niedergeschlagen (…) fröhlich (…) kann nie ohne einen verborgenen Haß des Gesetzes statt finden (…) das fröhliche Herz in Befolgung seiner Pflicht (…) das Gute auch *lieb gewonnen*”. RGV, AA 06: 23-4f. [↑](#footnote-ref-51)
52. “wahre Liebe (…) ein unentbehrliches Ergänzungstück der Unvollkommenheit der menschlichen Natur (…) denn was Einer nicht gern thut, daß thut er so kärglich, (…) daß aus diese als Triebfeder ohne den Beitritt jener nicht sehr viel zu rechnen sein möchte.” EAD, AA 8:337-8. [↑](#footnote-ref-52)
53. Grenberg 2014 tells a similar story but focuses on love for others and the law rather than the self. [↑](#footnote-ref-53)
54. If self-love of delight applies *only* to the moral part of ourselves, this would risk making the object of self-love overly generic. We want to love ourselves *in particular*, not only the part of ourselves that we share with every other moral agent. It seems Kant can adopt a more attractive view, according to which we love ourselves insofaras we are morally good but still love ourselves in our particularity, including all of the specific traits we have that contribute to our worth insofar as they are compatible with the moral law. Grenberg 2022 has recently developed a reading of the moral life as a happy or ‘eudaemonic’ life, in which the virtuous agent’s personal pursuits and goals become valuable to her as an expression of the value of her personhood. Grenberg does not appeal to self-love directly, but I believe our views can complement one another. [↑](#footnote-ref-54)
55. I cannot spell these implications out fully here; I merely mean to suggest that my account allows Kant’s views to connect to apparently true folk-psychological observations and to potentially enter into fruitful conversation with contemporary philosophy of love. [↑](#footnote-ref-55)