



# Intrinsically bad gratitude

Jörg Löschke

Received: 28 December 2019 / Accepted: 29 March 2021 / Published online: 26 April 2021  
© The Author(s) 2021

**Abstract** This paper discusses under which circumstances grateful responses can become intrinsically bad. It is argued that gratitude should be understood as an appropriate response to value and that it is subject to the so-called recursive account of intrinsic value, according to which appropriate responses to value are intrinsically good, and inappropriate responses to value are intrinsically bad. As a result, gratitude can become intrinsically bad in two cases: i) when gratitude has the wrong object, i.e., is a pro-response to an intrinsically bad act, ii) when the grateful response is grossly disproportionate, i.e., the intensity of the response does not match the goodness of the act that it is a response to.

**Keywords** Gratitude · Recursive account · Intrinsic value · Appropriate responses · Pro-attitudes

Gratitude has a good reputation. Parents teach their children grateful behaviour from an early age; psychologists advise us to keep gratitude journals because this promises more optimism, better health and a more positive outlook on one's life as a whole (Emmons & McCullough 2003). Many philosophers also hold gratitude in high regard. They argue that gratitude can contribute to solidarity and good societal relations (Fitzgerald 1998); that gratitude is a virtue (Meilaender 1984; Wellman 1999; Carr 2013; Manela 2019a); that people can be morally required to be grateful (Berger 1975; McConnell 1993); or that gratitude is an appropriate response to moral regard (Riedener *in press*). While these accounts differ in several respects, they all suggest that gratitude should be evaluated in a positive light, either due to its benefits or because of its moral status. And this is certainly in line with the way most people

---

J. Löschke (✉)

Philosophisches Seminar, Universität Zürich, Zollikerstr. 117, 8008 Zurich, Switzerland  
E-Mail: [joerg.loeschke@philos.uzh.ch](mailto:joerg.loeschke@philos.uzh.ch)



view gratitude: as something that is to be fostered and developed, and as something that is univocally good.

In this paper, I plan to pour some cold water on gratitude's good reputation. I do not aim to criticize any of the aforementioned views in detail. Nor do I deny that gratitude is often non-instrumentally good or that we should embrace, cultivate and teach our children grateful behaviour. But I will argue that gratitude is not univocally good. It also has a shadow side. Given certain circumstances, gratitude can cease to be good and can become intrinsically bad. And it is not necessarily or merely bad *for the agent* that she feels such gratitude. To say that gratitude is intrinsically bad is to say that the grateful response should not exist, or, to put things in somewhat grandiose terms, that the world would be better without the grateful response. I will argue that gratitude can be intrinsically bad under two circumstances. First, gratitude can become intrinsically bad if it has the wrong object – call this the content condition. Second, gratitude can be intrinsically bad if the intensity of the gratitude is too disproportionate to the value of the benefit for which the grateful agent is grateful – call this the proportionality condition.

Before I start my discussion, two clarifying notes are necessary. First, when I talk about the intrinsic badness of gratitude, I am talking about the intrinsic badness of individual grateful responses. I am not concerned with the question of whether gratitude is a virtue.<sup>1</sup> I will remain neutral on this question; as far as I can see, nothing in what follows depends on it. Thus, when I talk about gratitude, I mean individual responses of agents, rather than character traits or stable dispositions that give rise to such individual responses. Second, it is worth pointing out that I am only concerned with the evaluative status of certain occurrences of gratitude. I am not concerned with the question of under what circumstances an agent obtains a duty of gratitude.<sup>2</sup> How the evaluative and the deontic are connected is a complex question that I cannot discuss here.

## 1 What is Gratitude? General Remarks

I understand gratitude as a response to another person's beneficence, directed towards the grateful agent (or perhaps towards people with whom the grateful agent has a strong emotional connection, such as her children). A grateful response is thus a response to the (successful or perhaps unsuccessful) attempt by another party to

---

<sup>1</sup> For a discussion of this question, see Carr (2015).

<sup>2</sup> For a discussion of this question, see McConnell (1993).



benefit the grateful agent.<sup>3,4</sup> A grateful response has different parts, namely cognitive, affective, communicative and conative elements (Manela 2019b). The cognitive elements of gratitude are certain beliefs, in particular the belief that the beneficent act has occurred (and possibly the belief that the benefactor fulfilled certain conditions that must be met for gratitude to be appropriate).<sup>5</sup> Furthermore, these beliefs must be sufficiently persistent: if A benefits greatly from B's benevolent act but forgets about it the next day, then A certainly counts as ungrateful. The affective elements of gratitude consist in certain feelings, such as affection for the benefactor or goodwill toward the benefactor, which includes a desire that she fare well. To say that proper gratitude has a communicative dimension is to say that proper gratitude needs to be communicated, and that the forms of communication must be sensitive to the degree of benefit that the benefited person has received: for small benefits, a sincere 'Thank you' might be enough, whereas great benefits might require something else, such as writing a letter or giving flowers. Finally, gratitude includes conative elements: it implies certain modes of behaviour, including modes of treatment of the benefit that one receives.<sup>6</sup>

These four elements of a grateful response allow for a nuanced description both of gratitude and of forms of ingratitude. A person is fully grateful if her response to a beneficial act meets the standards for each element. If she does not meet the standards for one or more element(s), then she might either count as insufficiently grateful or as outright ungrateful (Manela 2019a; 2019b). Perhaps some of these elements are more important than others, so that a person who exhibits the right kind of feelings towards her benefactor but communicates her gratitude insufficiently counts as more grateful than a person who communicates her gratitude properly but lacks the necessary feelings towards her benefactor; I will leave this open at this point.

For analytic purposes, it is helpful to further qualify the content of the belief that is associated with a grateful response. The grateful agent believes that she has been benefitted (or that there has been an attempt to benefit her), and this means that she

<sup>3</sup> It is common to distinguish two kinds of gratitude, namely 'gratitude that' and 'gratitude to'. For this distinction, see, among others, Card (1988) or Manela (2016). In this paper, I will be concerned with 'gratitude to'. It is not clear whether 'gratitude that' really counts as gratitude, or that it should not simply be understood as a form of gladness or appreciation. See Manela (2019a).

<sup>4</sup> I understand beneficence as characterised (at least partly) by the intention of the beneficent agent. That is to say, the benefactor not only helps the grateful agent, but also intends to do so for the sake of the grateful agent. Thus, if A helps B with the intention to harm C, then this does not make gratitude on B's part appropriate – unless A has a mixed motivation and also acts for the sake of B. I thank an anonymous referee for pressing me on this point.

<sup>5</sup> In most accounts, such additional beliefs include the belief that the benefactor was motivated to benefit the benefitted person (see, among others, Berger 1975; Jecker 1989; Manela 2016). This is just the idea that gratitude is a response to beneficence. For some criticism, see Bardsley (2013). I will return to this point below.

<sup>6</sup> An anonymous referee has pointed out to me that it is also possible to understand the affective element and the conative element as elements of a piece. However, I think that distinguishing these aspects is helpful, as it allows for more nuanced descriptions of cases in which an agent fails in only one regard. For example, it allows us to say that an agent who feels goodwill toward her benefactor but has no inclination to act for her sake is less ungrateful than an agent who neither feels goodwill nor has the inclination to act for the sake of her benefactor.



believes the person she is grateful to has done something that is good for her (or at least she believes that he intended to do so). But she typically not only believes that the act was good for her, but also that it was good simpliciter. This is in line with the phenomenology of gratitude. An agent who has been benefitted by an act that she considers to be bad (say, someone helped her to get a job she really wants, but she later finds out that her benefactor murdered all competitors to enable her to get the job) will typically cease to be grateful. And if the grateful agent thinks that the beneficiary act is good overall but included intrinsically bad elements, then she will typically have a mixed reaction. If A saves B's life but must destroy a great work of art in the process, then B will typically to be grateful to A while at the same time mourning or feeling bad about the destruction of the work of art. A's destroying the work of art is not the object of B's gratitude, although it is part of a larger complex for which B is grateful. Thus, the intrinsically bad elements of the larger beneficent complex are typically not the object of the grateful response, and this suggests that gratitude includes not only the belief that one has been benefitted, but also the belief that the beneficiary act has been good simpliciter, or intrinsically good.<sup>7</sup>

To sum up: gratitude is a response to an act (or an attempted act) that the grateful person considers to be good for her as well as good simpliciter. And this response includes affective, communicative, and conative elements.<sup>8</sup> Since it is a response to something that the agent considers to be good, gratitude qualifies as a value response. And as a response to value, gratitude is subject to standards of appropriateness: value responses can be appropriate or inappropriate, and so can gratitude. This means that gratitude is subject to the same considerations as other kinds of value response, including the recursive account of value.

## 2 The Recursive Account of Value

The basic idea of the recursive account of value is that appropriate responses to intrinsic values are themselves intrinsically good. The recursive account is accepted by many value theorists,<sup>9</sup> and it is independently plausible. Its intuitive force can best be brought out when we consider examples.<sup>10</sup> Suppose that there are two worlds, W1 and W2. W1 contains suffering but no compassion, whereas W2 contains the same amount of suffering but also compassion. If we compare these worlds in terms

<sup>7</sup> For the purposes of this paper, I will use 'good simpliciter' and 'intrinsically good' interchangeably. Note that my formulation says that the grateful agent 'typically' considers the beneficiary act to be not only good for her, but also good simpliciter. Some philosophers think that there is no such thing as 'good simpliciter'. If it were a conceptual truth that the grateful agent considers the beneficent act not only good for her but also good simpliciter, then it would follow that such philosophers have no concept of gratitude, which strikes me as absurd. Hence, the claim that the grateful person considers the benefit to be good simpliciter should be understood as a psychological generalisation rather than a conceptual truth.

<sup>8</sup> To repeat, gratitude also has a cognitive element, but the response part of gratitude does not include the cognitive element. The response part of gratitude is a response to the content of the belief.

<sup>9</sup> See, for example, Nozick (1981); Chisholm (1986); Lemos (1994); Zimmerman (2001); Hurka (2001); Wedgwood (2009); Bradford (2013).

<sup>10</sup> For the following example, see also Hurka (2001).



of how good they are, then it seems plausible to maintain that W2 is the better world because it contains compassion for those who suffer. Now, compassion is arguably an appropriate response to the (dis-)value of suffering, and since the mere occurrence of compassion in W2 makes it the better world, independently of any further consequences (after all, the amount of suffering is the same), it is plausible to assume that it is intrinsically good to respond appropriately to value.

Another example that supports the recursive account is the case of sadistic pleasure. Sadistic pleasure seems intrinsically bad, even though pleasure as such might seem to be intrinsically good. The recursive account can explain the intuition that sadistic pleasure is intrinsically bad: being pleased by the pain of another person is not an appropriate response to the disvalue of that person's pain. Sadistic pleasure is an inappropriate response to value, and in the recursive account, this explains its intrinsic badness. So again, the recursive account yields the right result in this case.

Now, some readers might wonder *why* the recursive account holds. Why is it intrinsically good to respond appropriately to value? Different philosophers give different answers to this question. For example, Robert Nozick thinks that the recursive account follows from metaphysical considerations about intrinsic value. According to Nozick, intrinsic value consists in (or supervenes on) the structure of complex unity: whenever some entity is intrinsically good, this is because the entity brings disparate elements into a complex whole. And Nozick argues that appropriate responses to value also establish complex unity, namely unity between the valuable entity and the responding agent (Nozick 1981: 432). As a result, the recursive account simply follows from general considerations about intrinsic value. By contrast, Thomas Hurka thinks that the recursive account is a fundamental axiological fact for which no further explanation can be given (Hurka 2001). I do not have the space to discuss this question in detail, but I would like to make two remarks in favour of the recursive account and why we should accept it.

First, to see that the recursive account is very plausible, consider the possible views regarding the evaluative status of appropriate responses to value. Appropriate responses to value could either be intrinsically good, intrinsically bad, or evaluatively neutral. The view that it is intrinsically bad to respond appropriately to value seems obviously false. The view that appropriate (or inappropriate) responses to value are evaluatively neutral might be more plausible, but it does not seem very convincing either. It would imply that responding with compassion to the suffering of others and taking delight in their suffering are evaluatively on par, which is highly implausible. This leaves the recursive account by method of elimination. Of course, one could also argue that some appropriate value responses are intrinsically good while others are evaluatively neutral. But such a mixed view seems ad hoc; in any case, proponents of that view have the burden of proof to explain why appropriate responses to value have such different evaluative statuses. This, of course, does not explain why the recursive account holds, but it does suggest that we should accept it.

The second important consideration in favour of the recursive account is this. Value (or at least most instances of it) is not just out there in the world, independent of human interests or actions. As Joseph Raz puts it, '[...] intrinsic values are there to be engaged with by those who are of value in themselves. Their value is realised when those of value in themselves engage with them in the right way' (Raz 2001:



154). In other words, values need human activity for their realisation, even if the goodness of a valuable entity is mind-independent and not constituted by human valuing. And arguably that activity is to respond appropriately to value. If this is correct, then it helps to explain why the recursive account holds. If values depend for their (full) realisation on appropriate responses, then it is plausible to suppose that these appropriate responses are also good – after all, they realise value. And the value of appropriate value responses is not merely instrumental value in the sense that the value responses are contingent means to bring about good consequences. The responses constitute value; they do not merely cause it.

The question, then is: what counts as an appropriate or an inappropriate response to value? In a sense, it is not possible to give a general answer here that covers all cases, because the relevant standards of appropriateness depend on the valuable or disvaluable entity in question and different types of action or different kinds of emotional attitudes can be appropriate with regards to different valuable entities.<sup>11</sup> But it is possible to give a unified account of appropriate and inappropriate value responses on a sufficiently abstract level. Responses to intrinsic values can be subsumed under the umbrella terms of pro-attitudes and contra-attitudes.<sup>12</sup> As a result, we get four principles of recursion: (1) it is intrinsically good to respond with pro-attitudes to intrinsically good entities; (2) it is intrinsically bad to respond with contra-attitudes to intrinsically good entities; (3) it is intrinsically good to respond with contra-attitudes to intrinsically bad entities; and (4) it is intrinsically bad to respond with pro-attitudes to intrinsically bad attitudes.

But the direction of the attitudes is not the only thing that matters to the appropriateness of a response. Proportions also matter (Hurka 2001). In other words, the question is not merely whether an agent responds with a pro-attitude to an intrinsically good entity (or with a contra-attitude to an intrinsically bad entity). The strength of the attitude must also match the value or disvalue of the object of the response. If an attitude fails to meet this proportionality requirement, it can become inappropriate and hence intrinsically bad. The proportionality requirement might be more controversial than the requirement that value responses need to have the right direction, so let me make a few remarks here.

The first thing to point out is that an attitude does not become intrinsically bad if it is slightly too weak or slightly too strong. There are thresholds here, and while it might not be possible to give an exact formula to decide whether a given pro-attitude or contra-attitude satisfies the proportionality requirement, this does not mean that the proportionality requirement is false. We cannot measure exactly how good or bad an entity is, and we cannot measure exactly the strength of a pro-attitude or a contra-attitude either. This already shows that we cannot assign exact numbers to the exact threshold when a disproportionate attitude becomes inappropriate. And even if we could measure these things, I do not know of any way to fix the threshold at which a response with right direction becomes inappropriate, and hence intrinsically bad

<sup>11</sup> See Swanton (1995) for a discussion of different kinds of appropriate responses to value and why they are not reducible to one another.

<sup>12</sup> For a list of verbs that fall under these umbrella terms, see Nozick (1981: 429–430).



because it violates the proportionality requirement. Nevertheless, it is possible to identify cases where the proportionality requirement is clearly violated.

Consider first cases in which the attitude of the agent is too weak. Suppose that an agent is slightly annoyed by the fact that the Holocaust has happened. This is certainly an inappropriate response to the graveness of a moral catastrophe such as the Holocaust, and it is intuitively plausible to maintain that it is intrinsically bad to learn about the horrors of the Holocaust and respond merely with slight annoyance. But the verdict that slight annoyance is an inappropriate response to the Holocaust could not be explained if the direction of the attitude were all that mattered. After all, being annoyed by X is a contra-attitude towards X. Thus, proportionality matters when it comes to the appropriateness (and consequently intrinsic goodness or badness) of value responses.

At this point, one might object that this example does not support the proportionality requirement because annoyance is simply the wrong *kind* of attitude as a response to the Holocaust, and that the appropriate attitude would be something like moral outrage.<sup>13</sup> This might be taken to show that it is the kind of attitude that is relevant, not its strength. Now, it is certainly true that moral outrage is an appropriate response to the Holocaust. However, this does not show that the proportionality requirement is false. One important reason why outrage rather than annoyance is the appropriate response to the Holocaust is that outrage as such is a much stronger contra-attitude than annoyance is. Even the weakest form of outrage is a stronger contra-attitude than the strongest form of annoyance. Hence, the fact that outrage is the right kind of response to the Holocaust does not speak against the proportionality requirement, but actually supports it.

Furthermore, note that the proportionality requirement also applies in cases where the right kind of attitude is not at issue. Suppose that your mother has been in a car accident and you do not know whether she has been hurt. What would the appropriate response be upon your hearing that she was all right? Arguably, a feeling of relief and joy. But it would certainly be inappropriate to feel the level of joy that is appropriate upon finding a small amount of money while taking a walk, or to feel the level of relief that is appropriate when your favourite football team wins a fairly important game. And not only would such responses be inappropriate, they would also be intrinsically bad. Again, this intrinsic badness cannot be due to the wrong direction of the attitude. And it cannot be due to the wrong kind of attitude either. After all, it *is* appropriate to feel relief and joy in such a situation. It is just that the relief and joy in the inappropriate cases are too weak.

While these cases concern attitudes that are too weak, the proportionality requirement also applies when the response is too strong. Suppose that someone wrongs you in a minor way – say, she steals your parking space at the mall. While it might be appropriate to respond to the intrinsic badness of this act with mild resentment, it would certainly be inappropriate to resent her for weeks and weeks with extreme intensity. And it is also intuitively plausible that at some point such exaggerated resentment would become intrinsically bad, and that the world would be a better place without it.

---

<sup>13</sup> I thank an anonymous referee for pressing me on this point.



The most controversial cases are pro-attitudes towards intrinsically good entities that are much too strong.<sup>14</sup> Can such attitudes become intrinsically bad simply for being too strong? Upon reflection, I think that we should accept this possibility. Suppose that you hear that your mother has been in a car accident and that she is doing well, and that you feel the appropriate joy and relief. Suppose furthermore that, after hearing that your mother is well, you also hear that a painting of mediocre quality with very little (but some) aesthetic value which she was carrying in her car has not been damaged, and you respond with even greater relief and joy upon hearing this further news. This joy and relief seem so grossly inadequate that they become intrinsically bad. And again, this is not due to their wrong direction. Given that the painting has some aesthetic value, it *would* be appropriate to feel a little joy upon hearing that it has not been damaged. Or take the case of extreme selfishness (Hurka 2001). Extreme selfishness is an agent's response to her own good, and it is intrinsically bad. The best explanation for this intrinsic badness is that it is a pro-attitude to an intrinsically good entity that is too strong. Given that the agent's well-being is intrinsically good, it is appropriate for the agent to respond with a pro-attitude towards her own good. But in the case of extreme selfishness, this pro-attitude is far too strong. Extreme selfishness therefore also supports the view that pro-attitudes to intrinsically good entities can become intrinsically bad if they are (much) too strong.

If gratitude is plausibly thought of as an appropriate value response, and if the principles of recursion hold with regard to appropriate value responses, then it follows that the principles of recursion also apply to the case of gratitude. And this means that gratitude can be intrinsically bad. In particular, individual instances of gratitude can be intrinsically bad under two circumstances. They can be intrinsically bad if their object is intrinsically bad, and they can be intrinsically bad if the grateful response violates the proportionality requirement.

### 3 Intrinsically Bad Gratitude

I will refer to the two possible cases of intrinsically bad gratitude as the content condition (in which case gratitude has the wrong object) and the proportionality condition (in which case gratitude does not exemplify the right proportions between the goodness of its object and the strength of the response). Let me discuss these two possibilities in turn.<sup>15</sup>

<sup>14</sup> For a discussion of this possibility, see Hurka (2001). Hurka seems to be sceptical about this possibility.

<sup>15</sup> Some of the results of my discussion resemble the results of Tony Manela's (2019a) illuminating discussion of the virtue of gratitude and its associated vices. But there are important differences between Manela's account and mine. The first thing to notice is that Manela talks about virtues and vices, whereas I am concerned with the intrinsic goodness and badness of individual occurrences of gratitude. The second difference is that Manela does not emphasise the intrinsic badness of cases of gratitude in which the proportionality requirement is violated. For the most part, Manela seems to think that overgratitude is merely a sign for something else that is intrinsically bad; I will return to this below. A third difference is that Manela provides no unified explanation of the possible disvalue of various forms of inappropriate gratitude, whereas I hope to do so by applying the recursive account of value. I thank an anonymous referee for encouraging me to discuss the differences between my view and Manela's.



### 3.1 The Content Condition

Gratitude can be intrinsically bad if it has the wrong object. This happens if an agent is grateful for, and thus responds with a pro-attitude towards, something that is intrinsically bad. Different theories will give different accounts of what makes the object of the response intrinsically bad, depending on their specific axiology. For the purposes of this paper, I will understand the goodness or badness of the object of the grateful response in terms of the recursive account. Accordingly, the object of a grateful response (that is: the act that the agent is grateful for) is intrinsically good if it is an appropriate response to value, and the object of a grateful response is intrinsically bad if it is an inappropriate response to value. Several cases are possible here, and it is an advantage of the recursive framework that it provides a unified explanation of why gratitude is inappropriate in these cases.

The first and perhaps most obvious kind of case in which the content condition is violated is when an agent is grateful for an immoral act. Suppose that A is grateful to B for murdering a person that A finds annoying. Murdering a person because she is annoying is certainly an inappropriate response to that person's value and therefore intrinsically bad. Hence, A's gratitude is a pro-attitude to an act that is intrinsically bad, and the gratitude is intrinsically bad as a result. And this seems to be the right result. It is independently plausible that gratitude for deeply immoral acts is intrinsically bad, and that the world would be better without instances of such gratitude.

Now, following an influential paper by Justin D'Arms and Daniel Jacobson (2000), some readers might think that I am committing a moralistic fallacy here.<sup>16</sup> D'Arms and Jacobson convincingly argue that the fact it would be morally wrong to feel a certain emotion does not entail that the attitude is not fitting to its object, and that those who suggest otherwise commit a moralistic fallacy. The recursive account might seem to include such a moralistic fallacy, since it posits a close connection between inappropriateness and intrinsic badness. But this would be a misunderstanding. D'Arms and Jacobson are concerned with the fittingness of specific types of emotions and the question of whether the features that make those emotions fitting are in fact present. If, for example, the features that make a remark funny (whatever they are) are present, then the remark is funny, regardless of whether it would be wrong to laugh in the situation. By contrast, the recursive account is concerned with pro-attitudes as responses to intrinsically valuable entities. These are different things. For example, it can be an inappropriate value response to hold the pro-attitude of gratitude towards an intrinsically bad entity, even though it might be fitting to be grateful in the sense that the features that make gratitude fitting (such as the fact that the other person has benefitted you) are present.

In any case, it is important to note that not every act involving some element of wrongdoing is necessarily an inappropriate object of a grateful response. This would in fact be an unattractive form of moralism. Suppose that A benefits B greatly by telling a lie – say, B makes a mistake at work that will cost her job, and A lies to take the blame, thereby saving B from getting fired. In such a case, it seems

---

<sup>16</sup> I thank an anonymous referee for pressing me on this point.



wholly appropriate (and therefore intrinsically good) for B to be grateful to A, although telling a lie is at least *prima facie* morally wrong. How can we explain this intuition? Several lines of argument seem possible.

A first possible explanation is that A's lie in that situation is not morally wrong, all things considered. This might appear like the obviously right answer, but we should be careful here. After all, this line of argument presupposes that an act cannot be intrinsically bad if it is not morally wrong, all things considered. While this latter claim is not obviously false, it is nevertheless a position for which further arguments are needed.

A second possible explanation is that A's act is good overall, even though it includes an intrinsically bad element (in this case the telling of a lie). The difference between an act that includes murdering a person and an act that includes telling a lie is then that murdering a person is so intrinsically bad that an act that involves murdering a person can hardly be an overall good, whereas it is easier for an act that includes telling a lie to be an overall good as long as the benefit is great, given that the intrinsic badness of telling a lie can be outweighed more easily than the intrinsic badness of murdering a person. A possible problem with this explanation is that it suggests that appropriate responses always operate on the overall level and are responses to the overall goodness or badness of an entity. If the entity is good overall, then a pro-response is an appropriate response to that entity. But this makes it difficult to explain why it seems appropriate to respond with mixed reactions to entities that include both good and bad elements.

A third, and more promising, strategy to explain the intuition that it is appropriate for B to be grateful to A is thus to give a more nuanced description of the grateful response. We should understand gratitude not as a global response that occurs at the overall level and has an overall state of affairs as its object (such as the benefit plus the means that were necessary to provide the benefit). Rather, gratitude is a partial response: the object of the gratitude is solely the benefit that the other person has provided. The morally problematic means that were necessary to provide the benefit also call for appropriate responses, which might range from regret that they were necessary to outright horror that they were chosen. And depending on the appropriate strength of these other responses, they can outweigh or overshadow the grateful pro-attitude. Hence, the appropriate overall response to the provision of a benefit that includes murder is horror, since the appropriate strength of the contra-attitude to the horror completely outweighs and overshadows the appropriate pro-attitude to the benefit, whereas the appropriate overall response to the provision of a benefit that includes the telling of a lie might be gratitude, with a feeling of regret that telling the lie was necessary. This explanation not only seems to be the most promising one, but it also fits to the phenomenology of gratitude. It often seems appropriate to be grateful to another person even if you disapprove of some aspect of the benevolent act, and the idea that gratitude is not a global response explains such cases.

Cases in which gratitude has the wrong object because it is gratitude for an immoral thing are the more obvious cases, but there is also a second way in which a grateful response can violate the content condition: the object of the response itself can violate the proportionality requirement. According to the recursive account, responses to value can become inappropriate and therefore intrinsically bad if they

are either too weak or too strong, and therefore fail to match the value of their object. If the beneficent act for which the agent is grateful violates the proportionality requirement, then the agent is grateful for an intrinsically bad act.

Consider the following example. Suppose that Al is sitting in his favourite local coffee shop, working on a paper that he plans to submit soon. To take a break, Al wants to check his social media accounts to see whether one of his friends has posted a funny cat video that will amuse him for a couple of minutes. As soon as Al opens his browser, the wi-fi connection of the coffee shop collapses. Al cannot check his social media accounts, and this annoys him slightly. Beth is sitting at the next table and notices Al's slight discomfort. Beth is a very altruistic person, and as a result, she has a strong desire to help Al. She knows how to get the wi-fi running again, but the only way to do so involves sacrificing her life. Since Beth has a strong desire to help Al, she goes ahead and sacrifices her life to fix the wi-fi and enable Al to watch the cat videos.

What would the appropriate response to Beth's act be? Certainly not gratitude. We would not consider Al to be doing the appropriate thing when he shows up at Beth's funeral and expresses how grateful he is to Beth that she enabled him to watch cat videos. And this is not merely because the other people at the funeral are mourning their loss. Suppose that Beth was a loner and had no friends or loved ones. Even in such a situation, it would be inappropriate for Al to tell his friends how grateful he is to this stranger who sacrificed her life so that he could watch cat videos. It would instead be appropriate for Al to be shocked or to mourn Beth's sacrifice. Or, to put the point in more general terms: it would be appropriate for Al to have a contra-attitude towards Beth's act, rather than a pro-attitude. And since gratitude is a pro-response, gratitude would be inappropriate in such a case. And as an inappropriate value response, feeling gratitude toward Beth would be intrinsically bad.

But why would gratitude be an inappropriate response in this case? Many people think that it is morally permissible to commit suicide, and that it is morally permissible to sacrifice one's life for the sake of others. And if the self-other asymmetry holds, it is also permissible to forgo a greater good for oneself so that others can receive a lesser good. If all these claims are correct – and I will assume that they are – then it seems to follow that Beth acts morally permissibly when she sacrifices her life to enable Al to watch his cat videos. Hence, the intrinsic badness of Beth's sacrifice cannot stem from the moral wrongness of her act. A better explanation is that Beth's sacrifice is intrinsically bad because her act violates the proportionality requirement. Helping a person who is annoyed, or desiring that the person's annoyance should end, are appropriate responses to the disvalue of that person's annoyance. But these responses must be proportional to the disvalue of the annoyance. Given that being annoyed is arguably not of very great disvalue, Beth's reaction is clearly disproportionate: the value of helping Al to overcome his inconvenience does not warrant such a strong reaction as sacrificing her life. In fact, Beth's response is so disproportionate to the value of its object that her response becomes intrinsically bad, even though it is a pro-response to an intrinsically good thing.

To sum up, there are two ways in which a grateful response can be intrinsically bad due to violating the content condition: it can be gratitude for an immoral act,



or it can be gratitude for a morally permissible act that violates the proportionality requirement. But of course, this is not the only way in which gratitude can be intrinsically bad. The grateful response can violate the proportionality condition itself. This is what I will turn to now.

### 3.2 The Proportionality Condition

The proportionality condition is violated when the grateful response is either significantly too weak or significantly too strong, and gratitude can become intrinsically bad in each of these cases. It is difficult to determine exactly where the threshold is, but it is possible to identify several cases that clearly violate the proportionality condition. I will start with cases in which the grateful response is too weak.

It is important to keep in mind that a grateful response involves affective, conative and communicative elements. The grateful response can be too weak and hence violate the proportionality requirement in any of these dimensions, and the agent will appear ungrateful as a result. This choice of words is somewhat misleading, since it suggests that the agent demonstrates no grateful response at all. But arguably, ingratitude can also mean that the grateful response of a person is simply not strong enough. Suppose that Chris is trapped in a burning house, and Doris risks her life to save his. If Chris mumbles ‘I appreciate your help’ afterwards and goes his own way, then he certainly counts as ungrateful, even if we assume that he has the right affective and conative responses, and even though he communicates his gratitude to Doris. What makes Chris appear ungrateful is that he communicates his gratitude in an insufficient way. Hence, the mere fact that Chris’s response is too weak in one dimension can make him count as ungrateful and his response as inappropriate. And note that this is not an overall verdict about Chris’s response. Chris’s failure to show a sufficiently strong reaction in the communicative dimension cannot be compensated for by, say, him having a much stronger affective response than what would be called for. It is therefore necessary to discuss the three dimensions individually to see what it means for a grateful response to be insufficiently strong.

First, consider the affective element. It consists in positive feelings or goodwill towards the benefactor and the desire that the benefactor fare well, and it can be too weak in several ways. The strength of the positive feelings towards the benefactor must roughly match the degree to which the grateful person has been benefitted by the benefactor, and if these feelings are too weak, then this can violate the proportionality requirement. This does not mean that the grateful person must have overall positive feelings towards her benefactor. If her benefactor has intentionally harmed her in the past, it might very well be appropriate for the grateful person to have negative feelings towards her benefactor, but her negative feelings must at least become weaker to the extent that positive feelings towards her benefactor are appropriate. If her negative feelings towards her benefactor do not improve accordingly, then she violates the proportionality requirement.<sup>17</sup> The desire that her benefactor fare well can also be too weak, in two different ways. The first and obvious way is that the desire that the benefactor fare well is too weak. For example,

<sup>17</sup> See also Manela (2019b).



if Chris desires that Doris fare well after she has saved his life, but his desire is so weak that he does not really care, then this is certainly an inappropriate response to Doris's act, because even though Chris has a desire with the right content, that desire is not strong enough. A second way in which the desire that one's benefactor fare well can violate the proportionality requirement is if the desire has the appropriate strength, but the content of the desire does not match the value of the benefit. For example, Chris's desire that Doris fare well might be appropriately strong, but the content of the desire might be that Doris will fare well in the sense that she will win a card game against her parents-in-law, and nothing more. In such a case, Chris's grateful attitude also violates the proportionality requirement.

This brings me to the communicative element. The proportionality requirement can be violated here as well. This might appear odd – either a person communicates her gratitude or she does not, and there seems to be no room for considerations of proportionality. However, the example of Chris and Doris suggests otherwise. If Chris only mumbles a sign of appreciation, then he fails to communicate his gratitude to Doris sufficiently. Hence, the proportionality requirement seems to apply to the communicative dimension as well. And again, we can distinguish two possibilities here, as the proportionality requirement can be understood in a quantitative and in a qualitative sense. In a quantitative understanding, the grateful agent must express his gratitude often enough to satisfy the proportionality requirement, but this does not seem to be relevant sense in which the agent's communication of his gratitude must meet the proportionality requirement. Chris must express his gratitude appropriately, but that does not mean that he must do so over and over again. A more plausible understanding of the proportionality requirement is therefore a qualitative understanding. What matters is not how often the grateful agent expresses his gratitude, but how he expresses his gratitude. A possible problem here is that there is no intensity of communication that can match the intensity of the gratitude. But it is possible to account for proportionality in the communicative dimension by assuming a set of socially accepted symbolic forms for expressing one's gratitude that the members of a certain social group (and possibly also members of other social groups) can recognize as stronger or weaker expressions of gratitude. So, for example, simply saying 'thank you' counts as a weaker expression of gratitude than sending flowers or chocolate, and sending flowers and chocolate counts as a weaker expression of gratitude than writing a letter. And so on. The proportionality requirement is thus violated when an agent chooses to express her gratitude in such a way that her symbolic expression belongs to a lower class of socially recognized forms of expressing gratitude than the one that is called for.

What about the conative element? Again, it can be understood in two ways, either in terms of how strong the motivation for an action is, or in terms of the content of the action, and agents can violate the proportionality requirement in both ways. Suppose that Chris is motivated to help Doris, but his motivation is so weak that it is easily overridden – say, he notices that Doris suffered an injury while rescuing him and needs help to get to the doctor, and he is motivated to take her to the doctor, but instead watches every cute cat video as soon as a friend sends him a link. In such a case, Chris certainly counts as ungrateful, because his motivation to benefit Doris is not as strong as it should be. But Chris also counts as ungrateful if his motivation



to benefit Doris is as strong as it should be (he prioritises benefitting Doris over many other possible courses for action), but he is only motivated to benefit Doris in very minor ways, such as carrying one grocery bag to her apartment. If Doris has saved Chris's life and Chris is strongly motivated to carry one grocery bag for Doris but is not motivated to, say, give her a ride to the doctor when she needs one, then Chris has a grateful response, but that response is not strong enough. As a result, it appears to be intrinsically bad (simply for qualifying as an ungrateful response in spite of the fact that, strictly speaking, Chris is grateful – just not to the appropriate degree).

All cases that I have discussed so are cases where the agent has a grateful response, strictly speaking, but the response is too weak in an important dimension, therefore violates the proportionality requirement, and therefore becomes an inappropriate response to value and intrinsically bad as a result. But the proportionality requirement can also be violated when the grateful response is too strong. In such a case, the agent exemplifies affective, conative or communicative responses to the perceived benefit that are much stronger than the benefit warrants. As a result, such grateful responses can also become intrinsically bad.

I have mentioned before that it is controversial whether pro-attitudes towards intrinsically good entities can become intrinsically bad simply for being too strong, and the same applies to grateful responses that violate the proportionality requirement in being too strong. Again, it is admittedly controversial to maintain that such responses can become intrinsically bad. But I think upon reflection, we should accept that gratitude can become intrinsically bad for being too strong.

First, if we assume that pro-responses towards intrinsically good entities can be inappropriate and intrinsically bad for being too strong – and I have argued that this is a reasonable assumption – then the same holds for overgratitude. Of course, as in the case of other pro-attitudes towards intrinsically good entities, not every case of overgratitude is intrinsically bad. The grateful response must be much stronger than warranted for this to happen. However, if it is true that overshooting pro-attitudes can be intrinsically bad, and if gratitude counts as a pro-attitude, then it seems that the burden of proof is with those who argue that gratitude cannot be intrinsically bad for being too strong. And there are further considerations that support the claim that overshooting gratitude can be intrinsically bad.

As several authors have pointed out, too much gratitude can be bad insofar it exemplifies a lack of self-respect or servility on behalf of the grateful agent.<sup>18</sup> Furthermore, Tony Manela argues that overgratitude can lead to arrogance, as well as to an incorrect ranking of moral priorities (Manela 2019a). Now, Manela does not explicitly argue that overshooting gratitude as such is intrinsically bad. In fact, his point seems to be that overshooting gratitude is not intrinsically bad, but leads to things that are intrinsically bad (such as arrogance) or is a sign for something else that is intrinsically bad (such as servility or a lack of self-respect). But I think it is possible to make a stronger point here. Take the relation of servility and overgratitude as an example. Instead of understanding overgratitude as a sign of servility, it is

<sup>18</sup> See, for example, Card (1988). Manela (2019a) also maintains that overgratitude can be a vice, because it exemplifies the servility of the agent.

possible to understand overgratitude as one aspect of servility. According to this view, it is not the case that overgratitude is merely a sign for something else that is intrinsically bad. Instead, overgratitude at least partly constitutes an intrinsically bad thing such as servility. Thus, the intrinsically bad servility and the overgratitude are not different and independent things, but two aspects of the same intrinsically bad thing. If this is correct, then it provides an explanation of the intrinsic badness of (some cases of) overgratitude: overgratitude is a constituent part of an intrinsically bad thing, and as such, overgratitude is also intrinsically bad. This also explains why not all cases of overgratitude are intrinsically bad, but only those that are grossly disproportionate, for it is only such cases that also constitute servility.

#### 4 Implications of the Recursive Account

Applying the recursive account of value to the case of gratitude not only yields results that are intuitively plausible, but it also has some interesting implications. In closing, I will mention some of these implications. It will not be possible for me to discuss them in the detail they deserve, but they point to further promising lines of inquiry.

The first thing to notice is that there seems to be an asymmetry between displaying too much and too little gratitude. It is better to do the former. This, however, should come as no surprise if we apply the recursive account to the case of gratitude. Proponents of the recursive account already acknowledge such an asymmetry in their general framework. They agree that responses with the right direction can become intrinsically bad if they are too weak to a significant degree, but disagree on whether pro-responses can become intrinsically bad for being too strong. This demonstrates that many proponents of the recursive account already accept an asymmetry between pro-attitudes that are too weak and pro-attitudes that are too strong, and the case of overshooting gratitude and undershooting gratitude is just an application of this general phenomenon. Hence, if we accept the recursive account of value as a framework to discuss the phenomenon of gratitude, then this asymmetry is one implication of that general framework. Or at least, the asymmetry is compatible with the framework of the recursive account.

A second interesting point is this. Since a grateful response combines conative, affective and communicative elements, it is natural to suppose that it should be regarded holistically. And this suggests that a grateful response which violates the proportionality requirement in one of the aforementioned ways can be outweighed by a disproportionate response in another dimension. But this appears odd. Suppose that Adam has benefitted Christian in some moderate way, and that Christian feels no (or only very little) motivation to benefit Adam if the opportunity arises, but emphatically thanks Adam repeatedly. It seems plausible to maintain that Christian counts as ungrateful due to his lack of response in the conative dimension, regardless of his overreaction in the communicative dimension. This suggests that we should not understand the appropriate grateful response to a beneficiary act as a response with a certain strength, where this strength is a mere sum of the individual strengths of the conative, affective and communicative responses. And this leaves us with



two options. First, the relevant strength of the overall response could be a more complicated function of the strengths in the individual dimensions. The second option is more revisionary: perhaps we should reject the very idea that there is one grateful overall response, with the individual dimensions of response as parts of that response. This second option is revisionary because many authors who write about gratitude seem to assume that there is such a grateful overall response, and I have shared this assumption throughout this paper. But one possible upshot of my discussion is that perhaps we should reject this picture and should merely think of different grateful responses instead, so that an agent is not grateful or ingrateful *per se*, strictly speaking, but only conatively grateful (or ingrateful), affectively grateful (or ingrateful), and so on. While I do not have the space here to defend this second option in detail, I think it is the more plausible one, and it certainly would be an interesting implication of the recursive account when it is applied to the case of gratitude.

The last thing to mention is that the recursive account makes plausible claims about the comparable badness of content-related failures in gratitude and proportionality-related failures. *Ceteris paribus*, it is intrinsically worse to be grateful for an intrinsically bad act than it is to be insufficiently grateful for an intrinsically good act. This just follows from the very plausible assumption that the intrinsic badness of responding with an attitude of the wrong direction to an intrinsically good (or bad) entity cannot be outweighed by that (wrong) attitude having the right strength. Suppose that you respond with a pro-attitude to an intrinsically bad entity. How should the intensity of your attitude outweigh its wrong direction? It is certainly implausible to say that a contra-attitude to an intrinsically good object is intrinsically bad, but at least it has the right strength and is as strong on the negative side as the appropriate pro-attitude would have been, and this makes the contra-attitude better than it would otherwise have been. It is more plausible to maintain that contra-attitudes to intrinsically good objects are intrinsically bad, and the weaker they are, the better. But this means that the best possible attitude with the wrong direction has a strength of zero. In other words, proportionality cannot outweigh the wrong direction of a response – direction is more important than proportionality. As a result, content-related failures of grateful responses are worse than proportionality-related failures, all else being equal.

## 5 Conclusion

Gratitude is not always good. In fact, there are several ways in which gratitude can be intrinsically bad: it can have the wrong object, or the intensity of the conative, communicative and affective responses associated with a grateful response can fail to match the value of the benefit to the point that the grateful response becomes intrinsically bad. This follows if we understand gratitude as a value response, and apply the independently plausible recursive account of intrinsic value to it. To say that such cases of gratitude can be intrinsically bad is to say that the world as a whole would be better if these instances of gratitude did not occur.





Admittedly, this might sound a little grandiose. But I think we should take such possibilities seriously. We should not only teach children to be grateful; more importantly, we should teach them to be grateful for appropriate objects, and to an appropriate degree.

**Acknowledgements** This research has been funded by The Swiss National Science Foundation (SNSF) (Grant No. PP00P1\_176703). For helpful discussion and feedback, I am grateful to the audience of the conference on “The Shadow Side of Gratitude” (University of Erfurt), as well as an audience at the University of Zurich and two very helpful referees.

**Funding** Open access funding provided by University of Zurich

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bardsley, Karen. 2013. Mother Nature and the Mother of All Virtues: On the Rationality of Feeling Gratitude towards Nature. *Environmental Ethics* 35, 27–40.
- Berger, Fred. 1975. Gratitude. *Ethics* 85, 298–309.
- Bradford, Gwen. 2013. Evil Achievements and the Principle of Recursion. *Oxford Studies in Normative Ethics* 3, 79–97.
- Card, Claudia. 1988. Gratitude and Obligation. *American Philosophical Quarterly* 25, 115–127.
- Carr, David. 2013. Varieties of Gratitude. *Journal of Value Inquiry* 47, 17–28.
- Carr, David. 2015. Is gratitude a moral virtue? *Philosophical Studies* 172, 1474–1484.
- Chisholm, Roderick. 1986. *Brentano and Intrinsic Value*. Cambridge: Cambridge University Press.
- D’Arms, Justin & Daniel Jacobson. 2000. The Moralistic Fallacy: On the ‘Appropriateness’ of Emotions. *Philosophy and Phenomenological Research* 61, 65–90.
- Emmons, Robert & McCullough, Michael. 2003. Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Personality & Social Psychology* 88, 377–389.
- Fitzgerald, Patrick. 1998. Gratitude and Justice. *Ethics* 109, 119–153.
- Hurka, Thomas. 2001. *Virtue, Vice, and Value*. Oxford: Oxford University Press.
- Jecker, Nancy. 1989. Are Filial Duties Unfounded? *American Philosophical Quarterly* 26, 73–80.
- Lemos, Noah. 1994. *Intrinsic Value: Concept and Warrant*. Cambridge: Cambridge University Press.
- Manela, Tony. 2016. Gratitude and Appreciation. *American Philosophical Quarterly* 49, 55–66.
- Manela, Tony. 2019a. The Virtue of Gratitude and its Associated Vices. In: Roberts, Robert/Telech, Daniel (eds.): *The Moral Psychology of Gratitude*, 296–316. New York: Rowman & Littlefield.
- Manela, Tony. 2019b. Gratitude. In: Edward N. Zalta (ed.): *The Stanford Encyclopedia of Philosophy* (Fall 2019b Edition). (<https://plato.stanford.edu/archives/fall2019/entries/gratitude/>).
- McConnell, Terrance. 1993. *Gratitude*. Philadelphia: Temple University Press.
- Meilaender, Gilbert. 1984. *The Theory and Practice of Virtue*. Notre Dame, Ind.: University of Notre Dame Press.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge: Cambridge University Press.
- Raz, Joseph. 2001. *Value, Respect, and Attachment*. Cambridge: Cambridge University Press.
- Riedener, Stefan. Forthcoming. Beyond benefits: gratitude as a response to moral regard. *Inquiry*. <https://doi.org/10.1080/0020174X.2020.1820904>.
- Swanton, Christine. 1995. Profiles of the Virtues. *Pacific Philosophical Quarterly* 76, 47–72.
- Wedgwood, Ralph. 2009. Intrinsic Value and Reasons for Action. *Philosophical Issues* 19, 321–342.
- Wellman, Christopher. 1999. Gratitude as a Virtue. *Pacific Philosophical Quarterly* 80, 284–300.
- Zimmerman, Michael. 2001. *The Nature of Intrinsic Value*. Oxford: Rowman & Littlefield.

