THIS IS A DRAFT. PLEASE CITE THE PUBLISHED VERSION:
https://link.springer.com/article/10.1007/s11245-025-10238-9

Avoiding Toxic Charity in Argumentation

Katharina Stevens

1. Introduction

The interpersonal argumentative principle of charity requires that we interpret our interlocutor's arguments as presenting strong(er), rather than weak(er) (or no) reasons where plausible given the wording of the argument and relevant background information, even if this goes against our first, intuitive interpretation.[1] The principle is widely treated as a legitimate norm. We find it in public discourse, where even hostile opinion pieces pay it lip-service by offering a "charitable" interpretation of, say, a politician's words before proceeding with sharp criticism.[2] We see it in goal-statements of academic teaching, when professors say that they will teach students to treat philosophical arguments charitably even if they disagree (compare Lockard, 2022). We encounter it when we argue with our loved ones and they rebuke us for having treated them uncharitably. And it is present in almost every critical thinking textbook, even if it usually only

[1] This is my attempt at summarizing how the principle of charity is described in popular culture and in prominent critical thinking textbooks (Govier, 1988; MacDonald & Vaughn, 2016; Scriven, 1976; Watson, 2011). The interpersonal argumentative principle of charity is a sub-principle of the argumentative principle of charity more broadly understood, which would require that we interpret all arguments we come across charitably. However, I will not here address questions about how we ought to interpret arguments outside of interpersonal argumentation, such as when we privately contemplate arguments by historical figures like David Hume or Immanuel Kant, or when we engage with arguments that are offered to us by, e.g. AI. In addition, I will limit my discussion to what the argumentative principle of charity requires of arguers with respect to the *interpretation* of others' arguments. Some authors (Pruś & Sikora, 2023; Scriven, 1976) also discuss how arguers should select the objections they might offer to a certain argument (e.g. Scriven advises not to nitpick – i.e. to ignore minor flaws of the argument that could easily be repaired). I will here concentrate on charitable argument interpretation rather than charitable dialectical engagement with an argument. Finally, I should remark that I agree with Paglieri (2007), who argues that the seemingly automatic charity with which we understand enthymemes as complete, sense-making arguments is really just cognitive parsimony – not charity at all but resourcefulness in the allocation of cognitive resources. Therefore, I restrict my discussion of charity to instances where we have to resist our automatic interpretation of an argument in order to consider whether we should interpret it as strong(er) rather than weak(er) – charity, as I understand it, is not automatic. I also think that this is in line with typical textbook treatments: If charity was automatic, why would textbooks feel the need to remind students to engage in it?

[2] See, e.g., this article on Donald Trump's claim: "In four years you don't have to vote, ok? In four years don't vote, I don't care." https://www.msnbc.com/rachel-maddow-show/maddowblog/trump-tells-supporters-wont-vote-wins-rcna164019.

takes up around half a page (e.g. MacDonald & Vaughn, 2016; Warburton, 2007; Watson, 2011). Maybe more importantly, the interpersonal argumentative principle of charity is also supported by well-known moral and epistemic arguments that are routinely listed in those same critical thinking textbooks. There we can read that charity is required by fairness, since choosing to interpret arguments as weak undermines the arguer's ability to contribute meaningfully (MacDonald & Vaughn, 2016; Scriven, 1976). And we are told that charity promotes the accomplishment of epistemic goods since it supports the arguers' search for the strongest relevant reasons (MacDonald & Vaughn, 2016; Watson, 2011).

Given all this, you might be surprised to learn that a little research into the relevant literatures in argumentation theory and meta-philosophy reveals the principle's legitimacy as both morally and epistemically questionable. For every contribution that presents argumentative charity as morally required (e.g. Stevens, 2021), there is an equally convincing one detailing its morally toxic effects (e.g. Lockard, 2023). And for every author praising the epistemic advantages of charitable interpretation (e.g. Adler, 1981), there is one who bemoans how ill-advised charity poisons the epistemic community (e.g. Melamed, 2013).

I must admit that to me, the cumulative effect of reading this literature is disorienting. One can easily get the impression that charity is a kind of toxic duty; that it is a morally and epistemically grounded obligation the fulfillment of which creates moral and epistemic losses. I find this deeply unsatisfying. Therefore, my goal here is to determine why charity becomes toxic, where by "toxic" I mean that it does moral and epistemic harm instead of allowing arguers to fulfill moral duties and generate epistemic goods, as the arguments in its favor promise. And I want to find out whether arguers can comply with their reasons for argumentative charity while avoiding its toxic effects.

I begin by sorting through the epistemic and moral arguments for and against argumentative charity. Then I show that by distinguishing between three different *kinds* of charity, we can gain an insight into how it becomes toxic. The necessary distinction resolves an ambiguity in the requirement to interpret other's arguments as giving strong(er) reasons. On the one hand, an interpreter could aim at identifying a reading that makes the argument strong(er) from the interpreter's point of view. I will call this *egocentric* charity. On the other hand, they could aim at identifying an interpretation that allows them to understand how the argument presents a strong reason from the *arguer's* point of view. I call this *emic* charity. Finally, an interpreter could engage in *complex* charity. For this, they combine the two kinds of charity, for example by first employing emic charity to understand the arguer's point of view and then egocentric charity to find a way in which the arguer's point of view can be made meaningful for them.

I argue that egocentric and emic charity by themselves are toxic forms of charity because they undermine, rather than support, the effects that the arguments in favor of argumentative charity promise. But the problem with these forms of charity is not, as other authors have suggested, that they discourage interpreters from taking all the available evidence sufficiently into account (e.g. Govier, 1987) or that they make them assume the possibility of bad arguments away (Adler, 1996). Rather, the problem is that neither form of toxic charity helps arguers to both understand the perceived reasons their interlocutors are trying to communicate *and* take them seriously as potentially objective reasons. I close by arguing that, in contrast, complex charity assists with both. And while complex charity can sometimes result in epistemic and moral losses, this happens only when understanding and taking seriously another's reasons is itself inadvisable. Therefore complex charity is not toxic in the same way as egocentric and emic charity, even when it goes wrong.

2. The State of the Argument

2.1 Why Does Charity Need to Be Justified?

This paper grapples with the fact that the literature is divided about whether arguers should engage in charitable interpretation. For philosophers of language, the very existence of this problem may appear confusing. After all, in the philosophy of language, whether one ought to be charitable is not a decision made on moral or epistemic grounds. Rather, the principle of charity, especially Davidson's (1967, 1973) version, is presented as a necessary tool for understanding.

Davidson's principle of charity, according to which we interpret other's utterances in a way that optimises agreement, is conceptualized as necessary whenever there is any need of interpretation at all. Davidson introduces the principle as the key for resolving the dilemma of *radical interpretation*, where a speaker's utterances need to be interpreted while knowing *neither* what the speaker believes *nor* what they mean to say. Imagine, for example, a scenario in which the interpreter faces a speaker from a completely foreign culture, speaking an unknown language. Since we need to know what a person believes to determine what they mean, and we need to know what a person means to attribute beliefs to them, we are stuck. That is, we would be, unless we charitably assume that the speaker correctly (by our lights) perceives the world around them. This assumption makes radical interpretation possible, since it allows us to make assumptions about speaker's beliefs based on what we take to be true. According to Davidson, the principle of charity's requirement to optimise agreement also applies in non-radical interpretation. This is because the dilemma of needing to understand the utterance to determine what a person believes, but needing to know what they believe to understand the utterance, rears its head every time meaning isn't immediately clear.

As I said above, the *argumentative* principle of charity requires that arguments which can plausibly be understood in different ways be interpreted as offering strong(er) reasons, even if the interpreter's first inclination is to understand them as weak(er). Therefore, like in the philosophy of language, argumentative charity also requires the interpretation of utterances under the assumption that the speaker (arguer) is an apt epistemic agent.

However, the principle of charity in the philosophy of language aids the interpreter in their goal to understand and is therefore rather straight-forwardly *helpful.* By contrast*,* the interpersonal argumentative principle of charity places the interpreter into an uncomfortable motivational tension. This is because it addresses how arguments should be interpreted by the other involved arguers (from now on I will therefore call them interpreter/arguers). This introduces two complications. First, interpersonal argumentation usually takes place under conditions of (potential) disagreement or at least uncertainty. This means that the arguers need to treat their own and their interlocutor's points of view as at least potentially in question. With respect to the issue that they are arguing about, there is then no truth about the world that can simply be assumed. And second, the interpreter/arguer is engaged in argumentation with the person whose arguments they are interpreting. So they also face the task of critically evaluating them – which requires that they do *not* simply assume agreement (compare, e.g. Aikin, 2017).[3]

The result is that, on the one hand, the argumentative principle of charity demands that the interpreter/arguer exert effort towards improving the arguments' chances at being recognized, by them, as presenting a good reason. This is so because it requires them to take time and exert

[3] I think that the tension I describe here is what Lewinski (2012) has in mind when he rejects the idea that arguers may be required to interpret charitably – just that he assumes a stronger type of adversariality than I do (compare my (Stevens, 2021) criticism of Lewinski).

cognitive resources[4] to make sure that they interpret the argument as presenting strong(er) rather than weak(er) reasons instead of just going with their first, intuitive interpretation. But, on the other hand, because they are engaged in argumentation, they also must put up some resistance to this. So the principle can easily become a burden, especially because the effort that argumentative charity requires can be substantial: Arguments can be evaluated along several lines, including the acceptability of the premises, the strength of the inference, the openness to well-known objections etc. And interpretative charity can influence a whole range of decisions; how ambiguous terms are understood, how apparent contradictions in longer argumentative texts are resolved, whether missing premises are diagnosed and filled in, whether the passage in question even is an argument at all etc. (e.g. Govier, 1987; Scriven, 1976, p. 71 ff.; Stern, 2016; Tindale, Groarke, & Little, 2013, p. 140). Since, then, the argumentative principle of charity asks the interpreter/arguer to perform a sometimes very difficult task that potentially impedes them in their other duties, it is in need of justification.[5]

2.2 The Case for Argumentative Charity

Among the arguments that have been offered in support of argumentative charity, some appeal to the interpreter/arguer's self-interest. These prudential arguments tend to assume that the interpreter/arguer seeks to defeat the argument and, for example, counsel charity because it will save effort and embarrassment. After all, if an argument is criticized under a weak interpretation

[4] Even if, as in the dialectical account of Pruś and Sikora (2023), the interpreter achieves the charitable interpretation by engaging the arguer in dialogue about what they may have meant, this requires additional resources from the interpreter. This is so first, because the interpreter must be vigilant for occasions at which such dialogical engagement about the arguer's meaning is necessary (after all, we cannot ask after *every* argument whether we understood it correctly, we need to notice when our first, intuitive interpretation may be uncharitable). And second, it means that the interpreter must invest time and attention into the argumentative dialogue about the arguer's meaning.

[5] Compare also Paglieri (2007); Paglieri and Woods (2011), who critique the widespread notion that all interpretation of enthymemes ought to, or does, happen through charity by pointing out how costly in cognitive resources charity is. They suggest that it is instead cognitive parsimony that allows us to understand enthymemes as complete, functional arguments with relative ease.

when stronger ones are available, it will be easily repaired and new objections must be found to defeat it (Scriven, 1976; Watson, 2011).[6] However, most arguments for argumentative charity seek to show that the interpreter/arguer ought to interpret charitably independent of whether it serves their immediate interests.

Epistemic arguments advocate for argumentative charity because it is crucial for the pursuit of (one of) argumentation's essential goals, namely generating epistemically valuable outcomes like knowledge or understanding. This is because charitable interpretation makes the process of finding the best available reasons more reliable: On the one hand, it protects interpreters from missing out on a good reason just because it is hidden in a poorly formulated argument or because their cognitive biases make them predisposed to choose the weaker interpretation of an argument for a conclusion they reject (Pruś & Sikora, 2023; Stevens, 2021). It does this by ensuring that arguments are built up into their strongest form before they are exposed to objections, and that objections are not aimed at peripheral problems, thereby reducing the risk that premature criticism leads to the abandonment of fruitful ideas (e.g. Scriven, 1976). Additionally, during adversarial argument, being confronted with stronger arguments will motivate opponents to find better objections - thereby generating access to even more, and better, reasons (Adler, 1981; Melamed, 2013; Watson, 2011).[7] Since arguers ought to pursue epistemic goals during argumentation, they also ought to engage in argumentative charity.

[6] Of course, this assumes that the interpreter/arguer's opponent is quick-witted enough to repair the argument. If they are not, charity immediately becomes imprudent.

[7] Adler's (1981) version of this argument relies on Popperian falsificationism; he argues that submitting the argument to tests geared towards showing its falsity when it is in its strongest from is more useful from an epistemic point of view because it is more likely to reveal a flaw in the overall line of reasoning (I assume instead of a flaw merely in *this* formulation of the argument). Vedung (1983) points out that charity in argument interpretation makes most sense when the goal is to determine the epistemic merit of a position instead of evaluate the argument as made by a specific arguer.

Moral arguments present charitable interpretation as something that is owed to arguers and whose denial is unfair, harmful or disrespectful (MacDonald & Vaughn, 2016; Scriven, 1976). I have developed one of the more worked-out versions of this kind of argument (Stevens, 2021). I construct it from Govier's (1987) closely related conceptual argument. Govier points out that when argumentation is carried out in good faith, it inherently involves at least the attempt at mutually exchanging and evaluating reasons. Therefore, an interpreter/arguer who assumes that their interpersonal argument is mutually in good faith must also assume that their interlocutor is at least trying to give (strong) reasons and must, in turn, at least attempt to identify these reasons. I continue Govier's (1987) argument by pointing out that arguing with another person means at least pretending to pursue argumentation in good faith. After all, *openly* "arguing" without pursuing the mutual exchange and evaluation of reasons will not count as arguing (but instead as playing at arguing, having a bull-session etc.). *Not* engaging in charity, i.e. not exerting an appropriate amount of effort into identifying the interpreted arguer's reasons so that they can be evaluated and potentially integrated, is therefore a kind of deceptive free-riding on the otherwise cooperative enterprise of interpersonal argumentation. The uncharitable arguer accepts their interlocutor's willingness to engage seriously with *their* attempts at reason-giving but does not reciprocate. This is disrespectful to them as a reasonable being (I refer to this point as the deontological argument for charity) and it is harmful to them as a person whose practical and psychological well-being depends on their ability to participate meaningfully in reasons-based communication (I refer to this point as the consequential argument for charity). Charity is therefore morally required during interpersonal argumentation.[8]

[8] At least absent outweighing moral considerations against it. As both Govier (1987) and Lewinski (2012) point out, whether charity is morally justified depends on the context in which the argument takes place. It is quite possible that given the stakes involved, the reasons for charity are outweighed. This is why I formulate the conclusion of my (2021) paper carefully by saying that arguers have a *pro tanto* duty to be charitable.

2.3 Arguments Against Argumentative Charity

I think that the arguments in favor of argumentative charity are convincing. This is because I believe that, absent special circumstances, arguing ought to be oriented towards accomplishing epistemic goals and arguers owe each other a meaningful opportunity to make their reasons heard (Stevens, 2026). Further, I think that argumentation-based epistemic progress depends on the interpersonal identification and mutual acknowledgement of as many and as strong reasons as possible. And arguers can only make their reasons heard if their audiences are willing to invest effort into recognizing what they are trying to communicate, even against unfavourable first impressions. I find it plausible that requiring arguers to interpret arguments as presenting strong(er) rather than weak(er) (or no) reasons helps with both.

But this is not the whole story. The relevant literature shows that as much as charity can help generate epistemic goods, it can also bar access to them. And as much as charitable interpretation seems integral to enabling arguers to make their reasons heard, it can also be a tool for silencing them. I think that the main arguments against charity can be divided into two kinds:[9] First, those that show how charity can distort the interpreted arguments. And second, those that show how demands for charity can be weaponized to silence legitimate criticism.

The first worry, that charity may have a distorting effect on arguments, exists in both argumentation theory and in meta-philosophy.[10] In argumentation theory, it was introduced by

[9] There is at least one more kind of objection, namely that charity is too demanding on the arguer. Johnson (1980), for example, argues that some arguments, given the obvious unseriousness of the arguer, may not be worth the effort of a charitable interpretation. Similarly, Adler (1981) points out that charity can have epistemic costs if it leads to the use of valuable resources, which could be used better, for the improvement of arguments that are too unlikely to yield any kind of epistemic value. I will not include this objection in my discussion here since it is not one about the way that charity can lead to epistemic (or moral) losses through the way it influences the outcome of interpretation.

[10] It is important to note that, as I acknowledge in (Stevens, 2021), all interpretation will influence, and therefore "distort" the argument to some degree, simply because the interpreter is different than the arguer and has to understand the argument on their own terms. The kind of distortion that the authors discussed in the following paragraphs are concerned about is one that exceeds some threshold amount at which the argument as interpreted is so different from the argument as intended that the arguer would no longer recognize it as their own.

Johnson (1980) and then more fully formulated by Govier (1981, 1987). Her worry was that charity may erase the otherness of the other. This happens when it leads interpreters to generate a reconstruction of the argument that allows them to see it as presenting a strong reason, but that the arguer no longer recognizes as *their* argument (compare Melamed, 2013). Here, charity turns the arguer's argument into a vehicle for what the *interpreter/arguer* thinks are good reasons. I (Stevens, 2021) later called this distortive kind of charity "toxic" because it harms the pursuit of epistemic goods and generates moral wrongs; the opposite of what argumentative charity is supposed to do.

The pursuit of epistemic goods is harmed because, if what the interpreted arguer is trying to communicate really is a good reason, then access to this reason is impaired by the interpreter/arguer's influence. At best, it is replaced with access to another reason that the interpreter/arguer could have contributed themselves. At worst, instead of a genuinely new insight, the argument becomes a repetition of what was already commonly available (Govier, 1987). Further, even if the distorted argument is much better than the original, i.e. if the distortive charity constitutes a case of *iron-manning* (Aikin & Casey, 2016), the interpreter/arguer's manipulation still threatens to be epistemically costly and therefore toxic. First, the interpreted arguer misses out on the chance of recognizing their own argument's weaknesses through their interlocutor's criticism. Second, the interpreted arguer's interlocutors miss out on the insights they might have gained through serious critical engagement with the argument (compare Melamed, 2013).[11]

[11] Melamed (2013), writing about distortive charity in the interpretation of historical philosophers, describes what happens when interpreters reject the evidence for attributing a given view to an author because the view is "crazy" and the author a "great mind" i.e. too reasonable to hold a such view. The understanding that could have been gained about radically other ways to see the world is lost, and with it the epistemic gains that could have been generated by engaging seriously with it, including those that would have resulted from explaining exactly why it may be wrong or harmful.

Distortive charity also has significant moral costs because it robs the arguer of a meaningful chance to have their reasons heard as much as an uncharitable interpretation may. Apart from the disrespect that this expresses, it can also generate significant harm. In the most benign case, distortive charity iron-mans the argument in a way that the arguer themself prefers. Then the moral harm may be constrained to an unjust redistribution of credibility (compare Aikin & Casey, 2016; Bondy, 2010). But things can be much more serious than this. Where the interpreted arguer does not welcome the distorted version of their argument, this kind of charity effectively renders them silent. And if the arguer becomes aware of the interpretation and tries to correct it, since the resulting loss of voice is now an effect of the interpreter/arguer's effort to be charitable, mounting any defense against it risks generating resentment and is therefore even harder than defending against uncharitable interpretation.[12]

Lockard (2023) demonstrates how serious charity-induced loss of voice can be by describing its political dimension: Mainstream interpreters regularly use distortive charity to interpret arguments from minority speakers in a way that assimilates them to the mainstream. This allows them to reduce the discomfort of dealing with real difference while at least apparently acknowledging the interpreted minority arguer's contribution. Because the distorted version of the argument is easier accessible and less challenging to the majority, it has a good chance at eclipsing the original. If it does, it erases the challenge that the original argument could have posed to the established way of seeing the world.

The second worry, that the demand for charity can be weaponized to silence criticism, has its roots in what Lewinski (2012) has called the "paradox" of charity. The paradox is easily explained. Imagine two arguers, one of whom offers an argument while the other criticizes it.

[12] For an account of why it is so hard to defend against strawmen see Saussure (2018).

Interpreting either contribution charitably will make the other appear in a worse light – if, e.g. the initial argument is interpreted charitably, then the objection will appear at best as the result of a misunderstanding and at worst as that of straw manning (compare also Aikin & Casey, 2016).

Lewinski presents the paradox of charity as a reason for analysts to avoid using charity whenever interpretative uncertainty can be resolved by other means. When charity is necessary to settle on an interpretation, then it should be used only after careful evaluation of *who* should receive it, given the context. But Lockard's (2022) analysis of what happens when the *demand* for charity is used during interpersonal argumentation as a response to potentially legitimate objections reveals how dangerous the paradox is. Demanding argumentative charity can be used to reject an objection without having to show why it fails. Instead, it puts the onus on the *objector* to figure out how to let the argument survive their criticism and threatens them with appearing dim-witted or mean-spirited if they fail. This threatens the objector's owed chance to make their reasons heard, and generates epistemic and moral losses.[13]

3. Understanding Toxic Charity

3.1 From Egocentric to Emic Charity

I find the arguments against argumentative charity just as convincing as the arguments in favor of it. And again, it is because I endorse the idea that, absent special circumstances, argumentation should aim at epistemic goals and that arguers owe each other a meaningful opportunity to make their reasons heard. This is bad news, because taken together, these arguments produce an awkward tension. Now that argumentative charity seemingly supports *and* hinders epistemic goals and the

[13] Again, there is a political dimension: Lockard (2022) points out that charity is disproportionately demanded from minority voices when they criticize mainstream views, but rarely the other way around. This means that minority arguers are more often silenced through calls for charity which further cements power-imbalances.

fulfillment of moral requirements, what are arguers supposed to do? How does one reap charity's benefits while avoiding its toxic effects?

It may seem that the most straightforward solution is to clarify that argumentative charity is appropriate only in some contexts. Arguers should carefully consider contextual factors when determining whether to employ it. This is a popular answer among scholars who are involved in the charity debate and are aware of the considerations against argumentative charity.[14] And indeed, I believe that there is much wisdom in it. I agree with Govier (1987), for example, that using charity when confronted with arguers who clearly intend to deceive is a mistake. And I follow Lockard (2022) in warning of the repercussions when charity is selectively employed to shore up the viewpoints of the mainstream.

But I do not think that this answer is sufficient to help arguers walk the line between beneficial and toxic charity. To explain why, I first need to disambiguate the term "reasons" as it appears in the arguments in favor of charity from section 2. This will allow me to formulate a more detailed account of *how* argumentative charity can support the realization of epistemic goals and the fulfillment of moral requirements. And in turn, it will clarify what goes wrong when charity becomes toxic.

A reason is something that counts in favor of doing something (acting in some way or adopting a belief) for someone (e.g. Pinto, 2009; Scanlon, 2014). We can distinguish between *de facto* and *perceived* reasons. A *de facto* reason is something that does in fact count in favor of doing something for someone. By contrast, a *perceived* reason is something that *someone* perceives as a

[14] Lewinski (2012) and I (Stevens, 2021) both argue that whether argumentative charity should be employed and how strong it should be depends on argumentative dialogue the arguers are in. Govier (1987) and Johnson (1980) counsel arguers to employ charity only if argument or arguer meet certain threshold conditions having to do, for Govier, with the arguer's apparent intentions and for Johnson with the arguer's apparent effort and seriousness. Lockard (2022) suggests that even though she is deeply suspicious of charity, it can be employed in the right contexts to lift up marginalized voices.

de facto reason with some degree of confidence, however low (suspecting that something may be a reason counts for it to be a perceived reason). A person can be mistaken with respect to their perceived reasons; they may, for example, think that because they made a promise under threat, they ought to do what they promised. Then they have a *merely* perceived reason to do what they promised; a perceived reason that is not also a *de facto* reason.

Above, I introduced the argument that we ought to engage in argumentative charity to advance epistemic goals because it makes the process of finding the best available reasons more reliable. Presumably, this argument should be understood as referring to *de facto* reasons. Arguers (at least if they act in good faith) formulate arguments trying to present their perceived reasons, which they *take* to be *de facto* reasons with varying degrees of confidence.[15] Interpersonal arguing advances epistemic goals if it helps arguers increase the correspondence between their perceived and their *de facto* reasons. It does this by allowing arguers to bring their perceived reasons into intersubjective contact, revealing how the perceived reasons of different people confront and/or support each other. This, presumably, warrants a greater confidence in assigning the status of *de facto* reasons to those perceived reasons who survive the process and/or are affirmed by it - and denying it to those that don't survive.[16]

Now, argumentative charity requires that interpreters/arguers interpret the arguments of their interlocutors as presenting what *the interpreter/arguer* perceives to be a strong *de facto* reason (where plausible). Thereby, it assists the process of realizing epistemic goods insofar as it stops

[15] This description is admittedly oversimplified – arguers may also formulate arguments to present what they perceive as merely *potential* reasons for something, say to demonstrate that their interlocutor's perception of something as a reason is untenable. But I do not think that this oversimplification impacts my argument here.

[16] Different normative theories of argumentation will offer different accounts of how this works. Pragma-Dialectics, for example, subscribes to critical rationalism, according to which it is the *survival of confrontation*, i.e. attempts at falsification, that warrants greater confidence in the status of a perceived reason as a *de facto* reason (van Eemeren, 2018). By contrast, proponents of the epistemic theory of argumentation argue that the mutual support of perceived reasons can also warrant such increased confidence via *justification* (Biro & Siegel, 2008).

interpreters/arguers from mistakenly dismissing their interlocutors' *de facto* reasons as their *merely* perceived reasons because, e.g. they have been presented via poorly formulated or underdeveloped arguments. In adversarial settings, this also raises the quality of objections. That is the epistemic argument for argumentative charity.

We can clarify the moral argument for argumentative charity similarly. Above, I paraphrased the argument that we ought to engage in argumentative charity because we owe our interlocutors a meaningful opportunity to make their reasons heard. Presumably, to have one's reasons heard requires that one's perceived reasons are understood and given serious treatment as contenders for being *de facto* reasons.[17] Charity helps insofar as it prevents misunderstandings about what the arguer's perceived reasons are, which would also prevent them from being taken seriously. Such misunderstandings may for example be owed to the arguer's ineptitude in formulating their arguments and/or differences in the ways in which arguer and interpreter/arguer use terms etc.. This is the moral argument for argumentative charity.

Clarifying the arguments in favor of charity in this way shows why the instruction to use charity with discretion does not suffice to resolve the tension caused by the arguments for and against it. For imagine an arguer who attempts to communicate a perceived reason that is (let's stipulate) also a strong *de facto* reason. But they do so through an argument that initially appears very weak to their interpreter/arguer. The interpreter/arguer realizes that the argument could be interpreted differently so that it offers what they perceive to be a strong *de facto* reason. They do so, inadvertently distorting the argument. As a result, the arguer's attempt at making their reason heard is thwarted; their perceived reason is never treated as a serious contender for a *de facto*

[17] What it means to take a perceived reason seriously in this way will also vary depending on the normative theory of argumentation one subscribes to (compare footnote 10).

reason and never recognized as the strong *de facto* reason it is. The interpreter/arguer's charity was toxic in that it generated moral and epistemic losses.

It seems clear to me that the interpreter/arguer *should not* have used charity the way they did. But I do not think it equally clear that the situation would have been improved had the interpreter/arguer not employed charity *at all* and instead stuck to their initial interpretation of the argument as weak.[18] After all, the epistemic and moral reasons for charity that I just clarified still apply.[19]

Rather, I believe we would want to say that the interpreter/arguer should have interpreted the argument in a way that allows them to gain access to the (perceived and, by stipulation, *de facto*) reason of the *arguer*. That would have helped fulfill the duty of enabling the arguer to make their reasons heard, and it would have advanced epistemic goals. Since this would still have required resisting the initial interpretation of the argument as weak, doing what the principle of charity demands would still have been necessary – namely investing effort into seeing how the argument can be interpreted as presenting a strong(er), rather than weak(er) (or no) *de facto* reason. But the interpreter/arguer should have done this as a tool specifically aimed at accessing the

[18] I should note that a reader may wonder whether I am setting up a false dichotomy here; couldn't the interpreter just ask the arguer for clarification in order to make sure that they do not assume an interpretation that presents the argument as too weak, thereby avoiding both the employment of charity and going with their intuitive interpretation? I think that the interpreter could do that in some cases – but not always. As I already argued in (Stevens, 2021), an arguer may not be able to formulate the argument better than they did, they may well need the interpreter's help with that. In addition, I suspect that some effort towards charity can be required even just to notice that such a request for clarification is necessary. This is so because asking an arguer about the meaning of their argument first presupposes that the interpreter has noticed that there is interpretive insecurity and that the argument could be interpreted in different ways, some of which may present it as strong(er) than the interpreter thinks at first glance. I think that most cases where charity is lacking are cases where the interpreter does not even notice the possibility of different interpretations.

[19] Nor do I think does it make much sense to say that the interlocutor should have employed a charity that is *less strong* in the sense that they should have made sure to stay closer to the literal meaning of the words the arguer used. This is because I do not see how this would solve the problem that the charitable changes the interlocutor makes to their naïve interpretation, even if they are smaller, may lead them further away from accessing the reason the arguer attempted to communicate. (Compare the tendency, among argumentation theorists, to argue about how *strong* argumentative charity ought to be (Govier, 1981, 1987; Johnson, 1980; Lewinski, 2012; Scriven, 1976)

arguer's perceived reason. What they did instead was to identify, among their own perceived reasons, one that the argument could be made to present. In other words, the interpreter/arguer was engaging in a kind of *egocentric* charity aimed at making the argument fit into their own perception of how the world works and what constitutes a strong reason within it. And that was what made it toxic.

We can attempt to solve this problem by contrasting egocentric charity with another kind of charitable interpretation. This type aims to identify an understanding of the argument that makes accessible to the interpreter/arguer why the *arguer* perceives it to present a strong *de facto* reason, given the arguer's perception of how the world works and what constitutes a strong reason. An interpreter/arguer engaged in this kind of charitable interpretation would have to invest potentially significant resources into learning the arguer's view of the world as it is informed by their past experiences, emotional dispositions, normative commitments etc. (compare Stevens, 2021). Then they would have to attempt determining how the interpreted argument provides a strong *de facto* reason assuming this point of view.[20] Presumably, this would help with identifying the perceived reason the arguer was trying to communicate, again by working as a tonic against initial interpretations of the argument as weak owed to clumsy formulations or differing ways of using terms etc.

I want to call this kind of charity "emic" charity, borrowing the term from the emic/etic distinction as it is used in linguistics, anthropology and, more recently, work in intercultural argumentation (Mao, 2003; Mostowlansky & Rota, 2020; Pike, 1954; Tindale, 2021). The emic approach to researching a foreign culture aims at understanding how members of the culture see the world. An emic description of a social practice makes accessible how the practice makes sense

[20] In other words, they would have to engage in an exercise similar to what Gilbert (1997) suggests is required for what he calls "coalescent" argumentation.

from the point of view of those who engage in it. It is therefore only ever possible as part of a broad and appropriately detailed account of the entire culture and its associated normative and descriptive commitments. By contrast, the etic approach to researching foreign cultures starts with the development of a culture-neutral system of concepts. An etic description of a social practice then shows how it fits into a researcher-generated system of categories intended to be culture-non-specific.[21]

Interpreting an argument using emic charity means choosing to understand the argument as presenting what the interpreter/arguer perceives to be a strong(er), rather than a weak(er) (or no) *de facto* reason, just like interpreting it with egocentric charity would. But emic charity is more demanding than egocentric charity. This is because egocentric charity allows the interpreter/arguer to assume their own view of the world. Then they make the argument speak to this view, presenting what counts as a *de facto* reason within it. Emic charity, by contrast, requires that the interpreter/arguer learns to understand the way the arguer views the world. Then they determine how the argument fits into this view, offering what the interpreter/arguer perceives as a strong *de facto* reason within its context. As a result, emic charity aims at using the argument to identify the *arguer's* perceived reason because it acts as a tonic against misunderstandings. By contrast, egocentric charity aims at making the argument fit the interpreter/arguer's perceived reasons.

3.2 How Emic Charity Becomes Toxic

It is tempting to conclude that the argumentative principle of charity ought to demand emic charity during interpersonal argumentation.[22] After all, emic charity avoids the kind of toxicity that

[21] You may now think that I could have called egocentric charity "etic charity". But this would be a mistake: Etic research is not aimed at showing the foreign culture as it appears from the point of view of the researcher's home-culture. An etically oriented, Christian anthropologist would not, for example, seek to describe how Hindu practices can make sense from a Christian perspective.

[22] And in fact, I think this is what most argumentation theorists have done after the problem of argument-distortion became obvious in the literature. I think it is possible to interpret Scriven (1976), Vedung (1983) and maybe even

argumentation theorists and meta-philosophers have described for what we have now identified as egocentric charity. Since emic charity is oriented towards understanding the argument from the arguer's point of view, it lowers, not increases the risk that the interpreter/arguer will generate a version of the argument that is unrecognizable to the arguer. Therefore, it also does not bar access to the perceived reasons that the arguer tried to communicate. Further, since emic charity is not aimed at adjusting the argument to provide a strong(er) reason from the interpreter/arguer's point of view, demanding emic charity cannot block legitimate objections by forcing the interpreter/arguer to immunize the argument against them. It appears, then, that the epistemic and moral objections *against* argumentative charity do not apply to emic charity. Further, since emic charity helps the interpreter/arguer gain access to the arguer's perceived reasons, it also appears that the arguments *for* argumentative charity apply to it.[23]

Unfortunately, I think that these appearances are deceiving, and that emic charity can easily generate epistemic and moral losses of its own in a way similar to that of egocentric charity. But to show you why, I must disambiguate the term "reasons" once more and clarify the arguments in favor of argumentative charity even further. This will allow me to explain why emic charity can become toxic too.

Above, I said that a *de facto* reason is something that in fact counts in favor of doing something (to act or to believe) for someone. Now, I must add that not all *de facto* reasons are made equal. By following authors like, e.g. Vogelstein (2012), we can distinguish between

---

Adler (1981) as endorsing egocentric charity, but every newer, sophisticated account in the literature seems to go the emic way. Egocentric charity, in more recent treatments, seems to be reserved for textbooks (e.g. Watson, 2011). Apart from that, Paglieri (2007) seem to think there is a small set of circumstances in which egocentric charity is worthwhile, namely when the interpreter treats the arguer *solely* as an epistemic resource, a kind of inspiration-giver.

[23] Of course, emic charity comes with one rather obvious risk: If the arguer's worldview is odious – either absurd or deeply immoral – then immersing oneself in it may be a waste of time and resources. And in the worst case it can be epistemically risky, since opening up to something in order to understand it always means potentially being persuaded by it. I will address this worry in more detail in section 4.2.

subjective and *objective* reasons. A person has an *objective* reason to do something if something is in fact the case such that they ought to do it. A person has a *subjective* reason to do something if they have a belief such that, if the belief is true, they ought to do it. Vogelstein illustrates this with an example adapted from Parfit (2011): "While walking in a desert, you have angered a poisonous snake. You believe that running away will save your life, and believe nothing to suggest otherwise. As it turns out, however, you must stand still in order to save your life, as this snake will attack moving targets." In this example, you have a *subjective* reason to run, and an *objective* reason to stand still.

It is important to note that if someone has a subjective reason to do something, then they have more than a *merely perceived* reason to do it since they have a belief and, if that belief were true, then the corresponding fact would count in favor of them doing it (compare this to the example of the coerced promise, where the belief is true, but the fact does *not* count in favor of doing what they promised). The existence of their belief therefore is a *de facto* reason for them to do it.

Further, if a person has a subjective reason to do something then they *perceive* that they have an objective reason to do so. However, they may be mistaken about this because the beliefs that give them their subjective reason may be false. Then their subjective reason does not correspond to an objective reason.

Turning again to the epistemic argument in favor of charity, we can see that to work, it needs to be understood as saying that engaging in argumentative charity makes the process of finding *objective* reasons more reliable. After all, arguing advances epistemic goods if it generates such things as knowledge, understanding etc. All of these are dependent on increasing the correspondence of arguers' perceived *and* subjective reasons to objective reasons. Having *merely*

subjective reasons is not enough for generating knowledge, understanding etc., as the snake example above demonstrates.

Similarly, in order to work, the moral argument in favor of charity must be understood to show that charity helps with giving arguers a meaningful opportunity to have their reasons heard in the sense that their perceived reasons are understood and given serious treatment as contenders for being, not just *de facto* reasons, but *objective* reasons. After all, an arguer is owed a meaningful opportunity to make their reasons heard because they are owed an opportunity to contribute to a good-faith exchange and mutual evaluation of reasons. And we have just seen that this evaluation ought to aim at determining their status as *objective* reasons, since arguing ought to pursue generating epistemic goods. An interpreter/arguer who takes the perceived reasons of their interlocutor serious *only* as their subjective reasons treats the arguer as the object of inquiry rather than as a co-subject in the common enterprise of arguing about an issue.

At this point, we can understand why emic charity can easily become toxic. Emic charity demands only that the interpreter/arguer interpret arguments as presenting (strong) subjective reasons of the arguer. After all, to be emically charitable, the interpreter/arguer must seek understanding of the arguer's point of view, and then they must identify how, from this point of view, the argument offers such a reason. In other words, it requires the interpreter/arguer to identify why, if the arguer's beliefs *were true*, the argument *would* present an objective reason.

Unfortunately, interpreting an argument with the sole aim of discovering an arguer's *subjective* reasons during interpersonal argumentation can easily hinder, rather than help, the realization of epistemic goods and fulfillment of moral requirements. This is because, while emic charity may not result in a distortion of the argument, it does result in a distortion of how it is treated *as* an argument. An interpreter/arguer who uses emic charity does not treat the argument as

the potential representation of an objective reason that they may have to take on board as their own perceived reason. And they do not treat the arguer as a subject engaged in a reciprocal enterprise oriented towards producing epistemic goods. Instead, they treat the argument merely as an entry point for an inquiry into the arguer and their perceived reason. Emic charity, if used exclusively, *denies* the interpreted arguer the meaningful opportunity to make their reasons heard. And thereby it also *reduces* the chance that the (potential) objective reasons the arguer is attempting to present may help generate epistemic goods. Emic charity, by itself, is toxic charity.

Further, the demand for emic charity can be weaponized to block legitimate objections. This is because emic charity can generate its own version of Lewinski's (2012) paradox. Imagine an analyst interpreting the arguments of two arguers who are approaching an issue from different worldviews. Interpreting their exchange while engaging in emic charity with respect to one arguer means approaching the entire exchange from that arguer's worldview – which in turn risks rendering the other arguer's objections weak or even unintelligible. Again, we can rely on Lockard (2022) to understand the effects of this version of the paradox when emic charity is demanded of an interpreter/arguer: Such a demand forces the interpreter/arguer to enter the arguer's point of view and approach the argument exclusively from there. At best, this imposes the burden to translate their objections so that they make sense to the arguer, relieving them of having to engage in charitable interpretation themselves. At worst, namely if the arguer's worldview is seriously flawed and/or oppressive, it silences their objections altogether and locks the interpreter/arguer into what Lockard (2022, 2023), following Pohlhaus Jr. (2011), calls a hostile or oppressive world.

4. Avoiding Toxic Charity

4.1 Complex Charity

At this point, it may seem that there is no escape from the toxic effects of argumentative charity. Engaging in egocentric charity by interpreting arguments as presenting strong(er) reasons from the *interpreter/arguer's* point of view risks toxicity by distorting the argument. And engaging in emic charity by interpreting the argument as presenting strong reasons from the *arguer's* point of view *also* risks toxicity by distorting the treatment of the argument *as* an argument.

Still, I do not think we should give up on argumentative charity yet. So far we have only concentrated on what goes wrong. But we can also note what goes *right*: Using emic charity in interpretation helps the interpreter/arguer gain access to the perceived reasons of the arguer. And using egocentric charity helps them with finding the interpretation that is best at giving *them* a reason. That is, it helps them find the interpretation that shows the argument as presenting something that they may have to adopt as their own perceived reason because they recognize it as a serious contender for an objective reason.

In other words, each kind of charity does *part* of what the arguments in support of charity claim argumentative charity does. And each kind becomes toxic when it *fails* at doing or even undermines the other part. If argumentative charity is supposed to help generate epistemic goods and fulfill the moral requirements of interpersonal argumentation, then ideally, it supports both parts. It should help interpreter/arguers recognize the perceived reasons their interlocutors are trying to present *and* treat those perceived reasons as serious contenders for objective reasons. In other words, to reap the benefits of argumentative charity while avoiding its toxic effects, we need a kind of *complex* charity that unites in it the good parts of emic and of egocentric charity while avoiding their respective toxic effects.

Is this possible? I think so, but only by performing emic and egocentric charitable interpretation in sequence, so that each balances the other. One way this can look like is this: An interpreter/arguer can avoid the toxicity of egocentric charity by seeking insight into the arguer's perceived reason through engaging in emic charity first. Since this requires them to aim at interpreting the argument as presenting a strong subjective reason of the arguer, they must explore how the arguer sees the world. But, since they will eventually turn to egocentric charity, they must direct this exploration at identifying points of agreement or connection to their own view of the world. If found, these enable the interpreter/arguer to move into egocentric charity and attempt an interpretation of the argument as *also* presenting something that they can perceive to be a reason. Importantly, this attempt at egocentric charity is constrained by the already acquired understanding of the arguer's perceived reason and can therefore no longer result in a distortion of the argument. Or better, if a distortion of the argument were required to make the argument present something the interpreter/arguer can perceive as a reason, then the interpreter/arguer would now *notice.* The result of this (failed) attempt at egocentric charity is then the discovery of an objection. In sum, because it is constrained by the results of the preceding emic charitable interpretation, egocentric charity as part of complex charity does not become toxic. And it does not result in an immunization of the argument against legitimate objections. Quite the contrary, it now helps with their discovery.

Similarly, combining emic with egocentric charity into complex charity avoids the toxicity of pure emic charity. This is because the addition of egocentric charity ensures that the interpreter/arguer takes the argument seriously as a contender for presenting an objective reason. After all, by shifting to egocentric charity, the interpreter/arguer no longer treats the argument *merely* as a vehicle for gaining insight into the subjective reasons of the arguer. Instead, they now attempt to see how the argument could present something that can be a perceived reason for

themselves. And a perceived reason is a reason that the person in question perceives as a *de facto*, objective reason with some degree of certainty. So by moving to egocentric charity *after* engaging in emic charity, the interpreter/arguer attempts to see how the argument could present an objective reason and thereby avoids the toxicity of emic charity.[24]

But complex charity not only avoids the toxicity of pure egocentric or emic charity. It also fulfills the promises that the epistemic and the moral arguments in favor of argumentative charity make. Interpreting arguments with complex charity reliably (though, as I will soon admit, not always) supports the argumentative pursuit of epistemic goals. Emic charity prevents the premature dismissal of arguments based on clumsy formulation or misunderstandings. And the attempt at egocentric charity that follows ensures that the interpreter/arguer does their best at bringing the arguer's and their own perceived reasons into intersubjective contact in a way that validates the arguer's perceived reasons. If this attempt fails, and the interpreter/arguer discovers an objection instead, then this objection is now the result of a serious attempt at seeing the arguer's perceived reasons as contenders for objective reasons. In other words, it is the kind of objection the epistemic argument for argumentative charity promised.

For the same reasons, complex charity also helps arguers fulfill their moral responsibilities. The use of emic charity ensures that the interpreter/arguer makes a serious attempt at understanding the arguer's perceived reasons, and the subsequent use of (constrained) egocentric charity means that those perceived reasons are evaluated as serious contenders for being objective reasons. It seems that the principle of charity ought to require complex charity of arguers.

[24] I do not think that this way of combining the two types of charity into complex charity is necessarily the only way that works. One reviewer suggested that an interpreter/arguer may start with egocentric charity, and then offer the result to the arguer, whose response to it then helps them with engaging in emic charity and correct their first attempt at charitable interpretation. I suspect this could work too. In fact, it may be most useful to move back and forth between both kinds of charity. Here, I mostly intend to show that combining the two kinds of charity in sequence can help avoid charity's toxic effects.

4.2 Can Complex Charity Become Toxic?

It is never good to be overly optimistic. So instead of declaring that with complex charity, we have finally discovered a kind of argumentative charity that avoids toxicity, I ought to ask whether complex charity can become toxic too. I think the answer is both yes and no. Here is why:

The answer is yes, because it seems obvious that complex charity can generate epistemic losses. To see why, we must only realize that attempting to gain access to another's perceived reasons through understanding their view of the world always carries some epistemic risk. This is so simply because there is always the possibility that the interpreter/arguer may adopt the arguer's worldview or become convinced of their reasons even though the worldview is noxious and their reasons are *merely* perceived. Complex charity requires interpreter/arguers to take this risk because it requires emic charity.

Now, the risk is usually outweighed by potential gains. But not always. For example, Aikin (2019) describes a practice of the US American radical right to recruit adherents by attempting to cause a "non-rational, but nevertheless cognitively significant shift in perspective". The gestalt-shift is achieved by immersing them in the radical-right worldview so thoroughly that it simply replaces their original worldview. This is called taking a "red pill", after the red pill that, in the film *The Matrix*, allows takers to wake up from a computer-generated illusion and see the world as the dystopian nightmare that it is. I think it is clear that attempting to use complex charity when arguing with someone engaged in red-pilling poses an unacceptable epistemic risk for many, since the way in which access to the arguer's worldview is offered is specifically designed to cause non-rational adherence.

Admittedly, the practice of red-pilling is an extreme example. But it illustrates a more general insight. As Dutilh Novaes (2020) points out, arguers do and should consider whether they

trust their interlocutors when deciding whether and how to engage in interpersonal argumentation. An interpreter/arguer's willingness to engage in complex charity can be abused, especially if their rhetorical capacities are less sophisticated than those of the arguer. Where such abuse is successful, engaging in complex charity can generate epistemic losses.[25]

Complex charity can also generate moral losses. Arguably, in cases like the ones described above, the moral reasons interpreter/arguers have for engaging in complex charity may be outweighed by moral reasons having to do with the right to protect themselves (and potentially others) against morally relevant harms. The reasons for charity may still be there; even under conditions of severely diminished trust; and engaging another in argument *still* means at least pretending to engage them in an activity during which they will have a meaningful opportunity to make their reasons heard. So we may say that even then, interpreters/arguers have moral reasons against refusing complex charity. But we ought to acknowledge that where risks for epistemic losses are very high, especially if they will likely entail serious practical consequences, these reasons can be outweighed by reasons having to do with self-protection.

What's more, the argument that the right to self-protection may sometimes outweigh the moral reasons for giving others an opportunity for making their reasons heard can be extended beyond cases where charity risks epistemic losses. During my discussion of emic charity, I mentioned that Lockard (2022, 2023), following Pohlhaus Jr. (2011), worries that the demand for charity may force an interpreter/arguer into an oppressive world. If an arguer whose worldview is humiliating or deeply oppressive for the interpreter/arguer demands complex – and therefore also emic – charity, they also demand emotional and intellectual vulnerability to humiliation and oppression. This means risking serious, morally relevant harm. Depending on the extent of the

[25] Compare the discussions of the advantages and dangers of open mindedness, e.g. in Fantl (2018).

risked harm and the stakes involved in the argument, reasons having to do with the right to self-protection from this kind of harm may well outweigh moral reasons for charity. When this is the case, the refusal of complex charity is *at least* permissible. Consequently, where the demand for complex charity is backed with considerable power, demanding complex charity can be morally wrong.

In sum, engaging in complex charity can result in epistemic and moral losses and when these losses outweigh the gains, complex charity is epistemically and/or morally inadvisable (and demanding it can be morally wrong).[26] Still, I am hesitant to say that complex charity can become toxic in the same way as egocentric and emic charity. To explain why, I want to return one last time to the arguments in favor of argumentative charity.

The epistemic argument is built on the idea that epistemic goals are furthered when arguers gain access to and take serious each other's perceived reasons since this allows them to recognize how those reasons conflict with or support each other. And the moral argument is built on the idea that arguers ought to give each other meaningful opportunities to have their perceived reasons heard, i.e. understood and treated as serious contenders for objective reasons. Egocentric and emic

[26] We can identify a further situation in which engaging in complex charity may be inadvisable by considering the nature of deep disagreements (Fogelin, 2005). A deep disagreement exists when there is no common ground between arguers (or, in our case, arguer and arguer/interpreter) at all, and therefore also no foundation for emic charity to work from. Deep disagreements simply preclude mutual understanding. Complex charity can work when disagreement has depth – when there is *little* common ground – because its emic part can help expand the interpreter's understanding of the argument, thereby reducing the depth of the disagreement. But a deep disagreement, i.e. one of absolute depth, renders attempts at emic charity, and therefore complex charity, futile. All complex charity can do in cases like these is to waste resources and bring home, to arguer and interpreter alike, how incomprehensible they are to each other. So complex charity is not the right approach for dealing with absolutely deep disagreements. Depending on the circumstances under which the deep disagreement has been discovered, it may be better to engage in egocentric charity in order to see any reasons at all for which the interlocutor's position could be respected, or to break off arguing altogether. However, I do not think that absolutely deep disagreements are common. And I agree with Hitchcock (2020) that there are distinct moral dangers in calling something a deep disagreement prematurely. Therefore, I think that because arguers can usually not be sure how far down their disagreement goes – i.e. whether it is merely one of depth or absolutely deep – it is nonetheless often worthwhile to attempt complex charity first, if only in order to discover for certain that it is futile (compare Aikin, forthcoming).

charity become toxic because each undermines one part of what the epistemic and the moral argument promise, even while they help arguers achieve the other. But when complex charity goes wrong, it does not undermine either. Rather, it goes wrong exactly *because* it supports the interpreter/arguer in first understanding the arguer's perceived reasons and then taking them seriously as contenders for objective reasons. The problem is that doing this is either too epistemically risky or too morally harmful. In other words, in these situations it is a bad idea to do what proponents of argumentative charity have identified as the vehicle for argumentative epistemic gain and as morally owed during argumentation. In these cases, the reasons that speak against complex charity *also* speak against engaging in argumentation in good faith, full stop.

I do not mean to imply that whenever complex charity is inadvisable, people should not argue at all. I believe that there are situations in which the balance of reasons justifies engaging in argumentation in bad faith. An arguer may be strongly justified to believe that their potential interlocutor has mistaken and very harmful beliefs and that getting them to seriously consider arguments against those beliefs has a significant chance at generating improvement. At the same time, the arguer may be strongly justified in distrusting their potential interlocutor and in being worried that engaging in *good faith* argumentation would expose them to serious epistemic risks. Then it may be justified, all things considered, to engage the interlocutor in argumentation in bad faith, exploiting their willingness to seriously consider the arguer's perceived reasons while refusing to reciprocate. But I believe that in situations like these, we ought to be honest about what our arguer is doing. Namely, at the very least, they are committing a moral wrong against their interlocutor. Even if they are all-things-considered justified in doing this by more weighty moral reasons, they still have some reason for *regretting* not to engage in complex charity.

I suspect that this will be the case whenever complex charity is inadvisable. There will always be reason to regret that complex charity was not appropriate, all things considered, because arguing in good faith was not appropriate. This is not the case when egocentric and emic charity are toxic. And therefore, while I think that while complex charity can generate epistemic and moral losses, it cannot be toxic in the same sense that egocentric and emic charity can be. That is why I advocate understanding the argumentative principle of charity as a defeasible principle demanding complex charity.

Adler, J. E. (1981). Why Be Charitable? *Informal Logic, 4*(2). doi:10.22329/il.v4i2.2769
Adler, J. E. (1996). Charity, Interpretation, Fallacy. *Philosophy & Rhetoric, 29*(4), 329-343.
Aikin, S. F. (2017). Fallacy Theory, the Negativity Problem, and Minimal Dialectical Adversariality. *Cogency, 9*(1), 7-19.
Aikin, S. F. (2019). Deep Disagreement, the Dark Enlightenment, and the Rhetoric of the Red Pill. *Journal of applied philosophy, 36*(3), 420-435. doi:10.1111/japp.12331
Aikin, S. (forthcoming). Argumentation and Deep Disagreements. In S. Aikin, J. Casey, & K. Stevens (Eds.), *The Routledge Handbook of Argumentation Theory*: Routledge.
Aikin, S. F., & Casey, J. P. (2016). Straw Men, Iron Men, and Argumentative Virtue. *Topoi, 35*(2), 431-440. doi:10.1007/s11245-015-9308-5
Biro, J., & Siegel, H. (2008). In Defense of the Objective Epistemic Approach to Argumentation. *Informal Logic, 26*(1), 91. doi:10.22329/il.v26i1.432
Bondy, P. (2010). Argumentative Injustice. *Informal Logic, 30*(3). doi:10.22329/il.v30i3.3034
Davidson, D. (1967). Truth and Meaning. *Synthese (Dordrecht), 17*(3), 304-323. doi:10.1007/BF00485035
Davidson, D. (1973). Radical Interpretation. *Dialectica, 27*(3-4), 313-328. doi:10.1111/j.1746-8361.1973.tb00623.x
Dutilh Novaes, C. (2020). The Role of Trust in Argumentation. *Informal Logic, 40*(2), 205-236. doi:10.22329/il.v40i2.6328
Fantl, J. (2018). *The Limitations of the Open Mind*: Oxford University Press.
Fogelin, R. J. (2005). The Logic of Deep Disagreements. *Informal Logic, 25*(1), 3.
Gilbert, M. A. (1997). *Coalescent Argumentation*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
Govier, T. (1981). Uncharitable Thoughts About Charity. *Informal Logic, 4*(1). doi:10.22329/il.v4i1.2761
Govier, T. (1987). A New Approach to Charity. In *Problems in Argument Evaluation and Analysis* (pp. 133-158). Dordrecht: Foris.
Govier, T. (1988). *A practical study of argument* (2nd -- ed.). Belmont, Calif: Wadsworth Pub. Co.

Hitchcock, D. (2020). Deep Disagreement As Intellectual Colonialism. *OSSA Conference Archive, 12*. doi:https://scholar.uwindsor.ca/ossaarchive/OSSA12/Wednesday/28

Johnson, R. H. (1980). Charity Begins at Home. *Informal Logic, 3*(3). doi:10.22329/il.v3i3.2791

Lewinski, M. (2012). The Paradox of Charity. *Informal Logic, 32*(4), 403-439. doi:10.22329/il.v32i4.3620

Lewiński, M. (2012). The Paradox of Charity. *Informal Logic, 32*(4), 403-439. doi:10.22329/il.v32i4.3620

Lockard, C. A. (2022). *On the Uses and Misuses of Interpretive Charity in Philosophy.* ProQuest Dissertations & Theses, (Dissertation/Thesis)

Lockard, C. A. (2023). The Charitability Gap: Misuses of Interpretive Charity in Academic Philosophy. *Hypatia, 38*(1), 22-44. doi:10.1017/hyp.2023.1

MacDonald, C., & Vaughn, L. (2016). *The Power of Critical Thinking*. Don Mills: Oxford University Press.

Mao, L. (2003). Reflective Encounters: Illustrating Comparative Rhetoric. *Style (University Park, PA), 37*(4), 401-424.

Melamed, Y. (2013). Charitable interpretations and the political domestication of Spinoza, or, Benedict in the land of the secular imagination. In M. Lærke, J. E. H. Smith, & E. Schlisser (Eds.), *Philosophy and its history: Aims and methods in the study of early modern philosophy*. New York: Oxford University Press.

Mostowlansky, T., & Rota, A. (2020). Emic and Etic. In F. Stein, S. Lazar, M. Candea, H. Diemberger, J. Robbins, A. Sanchez, & R. Stasch (Eds.), *The Cambridge Encyclopedia of Anthropology*.

Paglieri, F. (2007). No More Charity, Please! Enthymematic Parsimony and the Pitfall of Benevolence. *OSSA Conference Archive, 114*. doi:https://scholar.uwindsor.ca/ossaarchive/OSSA7/papersandcommentaries/114?utm_source=scholar.uwindsor.ca%2Fossaarchive%2FOSSA7%2Fpapersandcommentaries%2F114&utm_medium=PDF&utm_campaign=PDFCoverPages

Paglieri, F., & Woods, J. (2011). Enthymematic parsimony. *Synthese (Dordrecht), 178*(3), 461-501. doi:10.1007/s11229-009-9652-3

Parfit, D. (2011). *On what matters*. New York;Oxford;: Oxford University Press.

Pike, K. (1954). *Language in Relation to a Unified Theory of the Structure of Human Behavior - Preliminary Edition*. Glendale, Calif.: Summer Institute of Linguistics.

Pinto, R. C. (2009). Argumentation and the Force of Reasons. *Informal Logic, 29*(3), 268-295. doi:10.22329/il.v29i3.2844

Pohlhaus Jr., G. (2011). Wrongful requests and strategic refusals to understand. In H. Grasswick (Ed.), *Feminist epistemology and philosophy of science: Power in knowledge*. New York: Springer.

Pruś, J., & Sikora, P. (2023). The Dialectical Principle of Charity: A Procedure for a Critical Discussion. *Argumentation, 37*(4), 577-600. doi:10.1007/s10503-023-09615-8

Saussure, L. d. (2018). The Straw Man Fallacy as a Prestige-Gaining Device. In *Argumentation and Language — Linguistic, Cognitive and Discursive Explorations* (pp. 171-190): Springer.

Scanlon, T. (2014). *Being realistic about reasons* (First ed.). Oxford: Oxford University Press.

Scriven, M. (1976). *Reasoning*. New York: McGraw Hill.

Stern, T. (2016). “Some Third Thing”: Nietzsche's Words and the Principle of Charity. *Journal of Nietzsche studies, 47*(2), 287-302. doi:10.5325/jnietstud.47.2.0287

Stevens, K. (2021). Charity for moral reasons? - A defense of the principle of charity in argumentation. *Argumentation and Advocacy, 57*(2), 67-84. doi:10.1080/10511431.2021.1897327

Stevens, K. (2026). *The Ethics of Argumentation*. New York: Routledge.

Tindale, C. W. (2021). *The anthropology of argument: cultural foundations of rhetoric and reason*. New York, NY: Routledge.

Tindale, C. W., Groarke, L., & Little, J. F. (2013). *Good Reasoning Matters!: A Constructive Approach to Critical Thinking, Fifth Edition*. Don Mills Ontario: Oxford University Press.

van Eemeren, F. H. (2018). *Argumentation Theory: A Pragma-Dialectical Perspective* (Vol. 33). Cham: Springer.

Vedung, E. (1983). Systematic Interpretation and the Principle of Charity. *Informal Logic, 5*(2), 21-22.

Vogelstein, E. (2012). Subjective Reasons. *Ethical theory and moral practice, 15*(2), 239-257. doi:10.1007/s10677-011-9286-7

Warburton, N. (2007). *Thinking from A to Z* (3rd ed.). London;New York;: Routledge.

Watson, J. C. (2011). *Critical thinking: an introduction to reasoning well*. London;New York;: Continuum.