What's so bad about fanaticism?

Paul Katsafanas

Draft of May 15, 2024

Forthcoming in *Synthese*

There is a longstanding debate about whether *fanaticism* can be good. We tend to imagine the fanatic as passionately committed to some cause or identity; as dogmatically, stubbornly resistant to compromising, attenuating, or rethinking that cause or identity; and as willing to resort to violence and coercion. We can certainly think of plenty of examples of vicious fanaticism: the insurrectionist who storms the Capitol; the incel; the jihadist; the white nationalist. Focusing on these violent, epistemically problematic, morally reprehensible individuals can make fanaticism seem terrible.

Philosophers have emphasized the terrible costs of fanaticism. In his *Philosophical Dictionary* entry on fanaticism, Voltaire tells us that fanaticism has "deluged England, Scotland, and Ireland with blood," for fanaticism has "inspired so many horrible cruelties." In Diderot's 1756 *Encyclopedia*, Alexandre Deleyre (1726–1797) writes that fanaticism "is blind and passionate zeal born of superstition, causing people to commit ridiculous, unjust, and cruel actions, not only without any shame and remorse, but also with a kind of joy and comfort" (Deleyre 2009). And that same risk seems alive today: when we view social discord fomented by fanatical groups, when we see the strife that seems to issue in part from fanatics, when we look at the horrific violence perpetrated by fanatics, fanaticism looks frightening.

But is fanaticism always bad? What about people who are passionately committed to a cause, are absolutely resistant to compromise, and are willing to carry out violent acts to achieve it, but who are directed at *morally good* ends?

One possible example is John Brown (1800–1859), who viewed it as his obligation to abolish slavery and famously claimed that he was willing to die—and kill—for this cause. Together with his sons, in 1856 he killed five supporters of slavery: pretending to be lost travelers, they knocked on peoples' doors, dragged them from their houses, interrogated them on their support for slavery, and then hacked them to death with swords. In 1859, he led a raid on the federal armory at Harpers Ferry, hoping to begin a violent uprising that would spread, overthrow the US government, and ultimately liberate slaves. 18 people were killed in that raid; many more were wounded.[1]

Brown was absolutely unwavering in his conviction that slavery was a great evil; he was willing to kill people and stoke violent uprisings in order to overthrow the US government; and his passion for his cause was unquestionable. But, of course, his cause was just: he correctly saw that slavery was an

[1] For accounts of Brown's life, see Oates (1984) and Reynolds (2006).

abomination and he turned to violence only because he saw it as the only hope for ending slavery. As he put it, in a note passed to his jailor on the day of his execution, "I, John Brown, am now quite certain that the crimes of this guilty land will never be purged away but by blood."[2]

What John Brown illustrates is that we can have people who are absolutely certain about their own moral convictions; who are willing to impose terrible costs on others so as to realize these convictions; who suffer, and in some cases die, in order to remain true to these convictions; who act with passionate intensity; but whose goals are laudable.

If John Brown is a fanatic—and there are indications that he is—then fanatics can be directed at morally good ends.[3] Indeed, some abolitionists willingly took that label upon themselves: in 1816, William Wilberforce (a leader of the British movement to abolish the slave trade) said: "They charge me with fanaticism. If to be feelingly alive to the sufferings of my fellow-creatures is to be a fanatic, I am one of the most incurable fanatics ever permitted to be at large."[4] And some philosophers make analogous points: Rousseau tells us that "fanaticism, although sanguinary and cruel, is nevertheless a grand and strong passion which elevates the heart of man, makes him despise death, and gives him a prodigious energy that need only be better directed to produce the most sublime virtues" (Rousseau 1979: 312).

Wilberforce and Rousseau are claiming that fanaticism is a mixed bag: while cruelty, violence, and social upheaval are in themselves bad, they can be directed in such a way that they are virtuous. John Brown is a possible example.

So, can fanaticism be good? Or is it always bad?

Recently, that debate has re-emerged in the philosophical literature. Quassim Cassam (2022b) and Paul Katsafanas (2019 and 2023a) have offered accounts of fanaticism that treat fanaticism as bad. Heather Battaly (2023) has argued that fanaticism is neutral: whether it is bad or good depends on several factors, including the context in which it occurs and the ends toward which it is directed. Olson (2007), Toscano (2017), and Steinberg (2023) also treat fanaticism as a morally complex state, one that is not necessarily bad.

My goal in this paper is to clarify what is at stake in these debates. At first, the debate about fanaticism can look merely terminological. After all, there's no agreed-upon definition of fanaticism,

---

[2] For a photograph of the original note, see https://commons.wikimedia.org/wiki/File:T-john-brown-last-prophecy.jpg.

[3] Nothing in my argument relies on determining whether Brown is a fanatic; if you don't think he is, you can substitute another example. But I do think there are indications that Brown was fanatical. I'll briefly review some evidence found in Griffen (2009), Oates (1984), and Reynolds (2006). As these sources make clear, Brown was often depicted as having messianic tendencies. His mother and some of his children had mental illnesses. Many of his contemporaries viewed him as "mad," "deluded," "fanatical," "crazed," and "maniacal" – these were all common newspaper descriptions of him. He was most commonly seen as monomaniacal, where, at the time, this was defined as the ability to behave normally in most situations but to be "irrational and obsessive" on certain subjects: "in pursuit of his or her obsession, the monomaniac could become implacable and ruthless, forsaking friends, family, even life itself in the headlong pursuit of a desired (and usually unrealistic) goal" (Griffen 2009: 372). Brown saw himself as specially chosen by God: when asked "Do you consider yourself an instrument in the hands of Providence?," he immediately replied "I do."

[4] William Wilberforce, address to the House of Commons, 19 June 1816.

so it might seem like this is just a debate about how to use the term: do we reserve it for problematic cases, or also extend it to good cases? But I argue that something more important is at issue. There's something common to the problematic and nonproblematic forms of extreme commitment that we see in these individuals. I'm going to argue that rather than trying to analyze fanaticism in a way that makes room for good fanatics, we should distinguish between an extremely robust and epistemically distinctive form of commitment—I'll call it *devotion*—and fanaticism.

Fanaticism is a pathological form of devotion. Fanaticism shares some of the good features of devotion, but adds to it a certain kind of vice. Nonetheless, fanaticism can have good effects in certain circumstances (just as other pathologies, such as having obsessive-compulsive disorder or being paranoid might have beneficial effects in local circumstances).

So I contend that philosophers including Rousseau and Battaly are wrong to claim that fanaticism is neutral or good. Fanaticism is bad. But there is something close to fanaticism—devotion—which is good.

The structure of the paper is as follows. Sections 1-3 introduce the concept of fanaticism; explore Cassam's and Battaly's accounts of fanaticism; and identify some problems with these accounts. Section 4 asks what is at stake in these debates about the goodness or badness of fanaticism. There, I argue that we should focus not on the normative status of the fanatic's object, but on the psychological profile of the fanatic. Section 5 identifies some respects in which fanaticism can be good. Section 6 introduces a distinction between *devotion,* which captures what is good in fanaticism, and *fanaticism,* which is a pathological form of devotion. With the distinction between devotion and fanaticism at hand, Sections 7-8 explain what is wrong with fanaticism. I conclude by explaining why the distinction between devotion and fanaticism is important.

1. What is fanaticism?

To figure out whether fanaticism is good or bad, we first need to know what fanaticism is. This point is in dispute—Katsafanas, Cassam, and Battaly offer different accounts of fanaticism, and when we turn to the early modern accounts of fanaticism we find even more divergence.[5]

Nonetheless, there is some shared ground. Katsafanas, Cassam, and Battaly—as well as many of the early moderns—agree that the fanatic has three features:

1. High degree of commitment: the fanatic has an extremely high degree of commitment to his object. He is wholeheartedly, unwaveringly committed to this object.[6]
2. High degree of certainty: the fanatic has rigid, unbending certainty about his object.
3. High degree of willingness to sacrifice: the fanatic is willing to impose costs on himself and others

[5] See Colas (1997), Crosson (2003), Katsafanas (2019), the essays in Katsafanas (2023b), Klein and La Vopa (1998), La Vopa (1997), Olsen (2007), Toscano 2010, and the essays in Townsend et al (2022).
[6] I use "object" in a very broad way. The fanatic's object can be an ideal, a goal, a cause, a person, a relationship, and so on.

Different philosophers analyze these features in different ways, and some of them add further features. Let's start with Cassam's account, which defines fanaticism in a way that makes it bad:

> Fanatics have *unwarranted* contempt for other people's ideals and interests, are willing to trample on those ideals and interests in pursuit of their own *perverted* ideals, and impose their ideals on others, by force if necessary. Fanatics are unwilling to think critically about their ideals because they regard them as indubitable. However, they are willing to sacrifice themselves and others in pursuit of their ideals. (Cassam 2022b: 133, emphasis added)

Cassam mentions a willingness to trample on other peoples' ideals and interests, as well as a willingness to sacrifice oneself and others; this corresponds to feature (3) above. He also mentions unwillingness to think critically about one's ideals; this mirrors (2). And he mentions having ideals, which incorporates (1). But he adds two further features: he says that the fanatic must have *unwarranted* contempt for other people's ideals and must have *perverted* ideals.

Cassam includes this condition because expressing contempt toward certain ideals (such as racist ideals, or oppressive ideals) can be warranted. And he includes the claim that the fanatic must have *perverted* ideals because he wants to say that certain abolitionists were not fanatics.[7] So, according to Cassam, a person who meets conditions (1)-(3), but has a non-perverted ideal, won't be a fanatic.

2. Problems with Cassam's account

So Cassam builds moral viciousness into his definition of fanaticism. The fanatic is vicious because his end is morally reprehensible.

However, this seems problematic for two reasons. First, it seems stipulative. Imagine two people who are as psychologically and behaviorally similar as can be, but one of whom is directed toward good ends and the other toward bad ends. Why think the one is a fanatic and the other isn't? It's not clear what additional work would be done by labeling the one a fanatic. If the only thing that differentiates the two is the nature of their ends, why not simply note that fact rather than introducing the terminology of fanaticism?

Second, in ordinary discourse the term "fanatic" doesn't seem to condemn the person's *object*. Calling someone fanatical involves making a negative judgment about the person (i.e. their psychology), but does not involve condemning their object. One indication of this is that people can be fanatical or non-fanatical toward the same object: even if we pick a morally problematic end, we can distinguish between the person who is fanatical about that end and the person who relates to it in a different way.

So why would we want our account of fanaticism to claim that the fanatic's end must be vicious/problematic? Go back to John Brown. His ideal isn't vicious; it's meritorious. If the only thing that differentiates him from the fanatic is that their ideals differ in content—one being meritorious, the other reprehensible—then it seems strange to label the one a fanatic and the other a non-fanatic.

---

[7] In particular, Garrison: Cassam (2022a) argues that William Lloyd Garrison wasn't closed-minded, that he was able to think critically and strategically, and so on. Garrison was a radical, seeking sweeping change, but not a fanatic.

We can bolster this point by considering analogous cases. We don't have different words for *obsession* directed at good things and at bad things. We don't have different words for *stubbornness* directed at good and bad ends. We don't have different words for *arrogance* about good and bad matters. We don't have different words for *cruelty* directed toward good and bad ideals. And so on. And that's because these concepts pick out facts about the *person*, rather than facts about the contents of the person's goal or ideal. Just so, I suggest, with fanaticism.

3. A broader account

Let's consider an attempt to develop a broader account of fanaticism. Heather Battaly offers such an account:

> "I will be suggesting that their [Cassam's and Katsafanas'] analyses are too narrow and fail to capture fanaticism in its more mundane forms, including fanaticism about activities (e.g., gaming, fitness, bird-watching), topics (e.g., bacteriology), sports teams (e.g., the New York Yankees), musicians (e.g., Beyoncé), celebrities (e.g., Justin Bieber), and products (e.g., iPhones, beauty products). Zooming out and approaching the subject from a different vantage point can help us identify conditions that are not only necessary for violent forms of political and religious fanaticism, but for fanaticism more generally." (Battaly 2023: 6)

As this quotation indicates, Battaly's account would include not only people with non-perverted ideals, but a huge range of people who are obsessively committed to activities, hobbies, teams, and so on.

According to Battaly, to qualify as a fanatic about some object, a person must meet the following three conditions:

- "F1. Be disposed to care so strongly about the object in question that one's motivation to pursue it outweighs one's other strong motivations and objectively counts as a passionate obsession.
- F2. Be disposed to focus on opportunities to pursue the object in question at the expense of opportunities to pursue other things that one is strongly motivated to pursue, where one's focus also objectively counts as obsessive.
- F3. Be disposed to act in pursuit of the object in question even at the expense of acting in pursuit of other things that one is strongly motivated to pursue, where one's behavior also objectively counts as obsessive." (Battaly 2023: 13)

In short, the fanatic is obsessively passionate, obsessively focused, and obsessively motivated toward an object. The object can be morally positive (the abolitionist), morally neutral (the baseball fan), or morally pernicious (the member of a hate group).

While Battaly is correct to dissociate fanaticism from negative ideals, there are some problems with the account that she offers. First, Battaly's account does not distinguish fanaticism from mere obsession or compulsion. This is built into her conditions: the fanatic is defined as obsessively passionate, obsessively focused, and obsessively motivated. But it's not clear that obsession always qualifies as fanatical. Imagine two different cases. First, imagine a person who is starving and in

desperate need of food. Such a person is likely to qualify as meeting F1-F3, but it would be odd to count them as fanatical. Second, imagine a person who is extremely devoted toward their child. They care so deeply about the child that their child-directed motives outweigh all other motivations; they are obsessively devoted to focusing on their child, with most of their time and energy spent on this; and they are obsessively directed toward pursuing their child's welfare, where this comes at the expense of other strong motivations (such as career, friends, etc.). I think this is a relatively common occurrence and not one that we would typically describe as fanatical. It's obsessive; it might be in certain respects cramped or blinkered; but it does not strike me as fanatical.

Second, Battaly's account seems too broad. It's true that we use the term "fanatic" to refer to passionate sports fans, people who are obsessive about celebrities, and so on, but that seems like a derivative or metaphorical use. If we take, as our paradigms of fanaticism, the violent, coercive, epistemically problematic, socially disruptive individuals mentioned at the outset of this essay, they seem strikingly different than the person who is innocuously obsessive about the Red Sox or Taylor Swift or stamp collecting. We should avoid blurring the very important distinctions between these people.

While I criticized Cassam's account for being too condemnatory, Battaly's goes too far in the other direction: it is too neutral. For, in a limited way, I agree with Cassam: whatever fanaticism is, it should be something bad. Cassam secures the badness of fanaticism by treating the psychological profile of the agent as fine and the ideal as perverted. Below, I am going to argue that fanaticism's badness lies not in the ideal but in the psychological profile.

4. Taking stock: what's at issue in these disputes?

As the above discussions demonstrate, there are different ways that fanaticism could be bad. Cassam claims that fanaticism is bad because the fanatic's object is bad: fanaticism is defined in a way that requires a morally perverted ideal. Battaly disagrees, claiming that fanaticism is neutral because the fanatic's object can be either good, bad, or neutral. Both of these views agree that if fanaticism is bad, the badness of fanaticism will lie in the badness of the fanatic's object. I argue that this is a mistake. For there is another possibility: fanaticism's badness lies not in the object, but in the person.

In short: I will argue that fanaticism is a vice. Like other vices, it can be present in a person who is directed at good ends. And it can even be connected to individual or social flourishing: there can be conditions under which it's all things considered better, either individually or socially, to be a fanatic.

To make that point, I first need to offer some clarifications. Below, I'll consider what's supposed to be good about fanaticism—what is it, about fanaticism, that tempts some thinkers to treat it as praiseworthy? I'll argue that there are certain forms of closed-minded, gritty, dedicated commitments that are praiseworthy, and those forms of commitment show up in fanaticism. However, I'll argue that we should tease apart a neutral notion—a form of closed-minded, gritty, dedicated commitment, which I call devotion—and use it to pick out those features. We should reserve fanaticism for a pathological form of devotion, which couples devotion with a problematic stance. Insofar as devotion to some end is good, fanatical devotion to the end will also be in certain respects good. But it will still be a pathology, involving a problem with the person's psychology.

Distinguishing fanaticism and devotion, while treating fanaticism as a vicious form of devotion, clarifies what is at stake in the debates about whether fanaticism is bad. There are three important points.

First, in ordinary discourse we typically treat fanaticism as bad. There should be at least a minor presumption in favor of accounts that match customary usage. Treating fanaticism as a vicious form of devotion enables us to do so. Moreover, it enables us to do so while still recognizing that many praiseworthy forms of devotion resemble (but are not identical to) fanaticism.

Second, one underlying disagreement in these disputes about the badness of fanaticism is whether every case of resolute, doubt-free dedication to costly ends constitutes fanaticism. Many forms of political and social activism involve the three core features of fanaticism (high degree of commitment, high degree of certainty, high degree of willingness to sacrifice). Consider someone like Nelson Mandela, who devoted himself in a wholehearted, unbending manner to his cause and was willing to endure extreme costs to sustain it. More generally, the champion of civil rights, the environmental activist, the person struggling against oppression might all display these three features. So, too, many important personal relationships involve these features. For example, parents often have high degrees of commitment, certainty, and willingness to sacrifice for their children. Some accounts of fanaticism, like Battaly's, will classify many of these people as fanatical; others, like Cassam's, will say that these people avoid fanaticism only to the extent that their ends are not vicious. But if an account of fanaticism ends up lumping these dedicated activists and committed parents together with fanatics, it misses an opportunity to tease apart what is admirable in these figures—the dedication, commitment, and wholeheartedness—from what is vicious. So one thing that is at stake in the debates about fanaticism is whether we end up classifying all forms of extreme, costly, wholehearted dedication to ends as fanaticism.

Third, some accounts of fanaticism will enable us to identify conditions that might otherwise be overlooked or misconstrued. Philosophers in the early modern period were interested in fanaticism in part because they saw it as a *distinctive* source of social discord.[8] We can ask, of any account of fanaticism, whether it identifies something distinctive, something for which we do not already possess other concepts. Assimilating fanaticism to mere obsession, as Battaly does, groups together innocuous forms of extreme commitment with horrific ones; restricting fanaticism to morally objectionable ends, as Cassam does, makes the concept somewhat superfluous (as we could simply point to the objectionable ends). But if there is a problematic way of manifesting extreme commitment, which arises from a distinctive psychological profile and is linked to social pathologies, then we risk occluding it by mischaracterizing fanaticism.

The debates about whether fanaticism is bad, then, are not mere terminological disputes. A failure to characterize fanaticism properly risks incorrectly assimilating innocuous and praiseworthy forms of devotion to problematic ones, and thereby obscuring a distinctive vice.

5. What's good about fanaticism?

So, what might be good about fanaticism?

[8] See Colas 1997, Goldsmith 2022, Katsafanas 2019, Katsafanas 2023b, La Vopa 1997, Toscano 2010, and Townsend et al 2022.

First, fanaticism can promote or sustain good ideals, goals, or states of affairs. Many meaningful ideals, goals, and projects require long-term, sustained effort in the face of challenges, setbacks, and resistance: think of political goals, social goals, personal relationships, and so on. Fanatics are dedicated: they are willing to commit to their ends despite adversity and to persevere in the face of setbacks. In light of this, fanaticism can have good consequences. Take the abolitionist movement: this was a generations-long struggle that required immense sacrifice. It's reasonable to assume that if there hadn't been people like John Brown—people who were willing to suffer, and to make others suffer, for the sake of this ideal—then it would not have been achieved, at least not in the same amount of time. If we assume that John Brown was a fanatic, and that John Brown helped to bring about the end of slavery, then this instance of fanaticism had a good outcome. (If you don't like that example, invent a hypothetical example in which fanaticism does in fact do this. There's no reason to deny that this can happen.) So fanaticism involves an extremely robust, resilient form of commitment. And we sometimes need robust commitment to achieve important goals.

Is this enough to show that fanaticism is good? Not really—not in any substantial sense. We already know that bad things can have good effects. Pick any vice (envy, greed, etc.), and you can find circumstances in which it generates good effects. Pick any horrific way of living one's life, and you can find circumstances in which that's a better way to live than its alternatives (e.g., being a cruel, sadistic person who serves in a group fighting for a good cause, where the sadism is linked to the group's success). So it's not particularly surprising that fanaticism can promote or sustain good states of affairs.

Second, fanaticism can gratify certain existential needs. A number of philosophers and psychologists have argued that we want to feel that our lives are meaningful, or that we have a purpose, or that what we are doing matters.[9] Fanaticism can answer these existential needs. Insofar as you wholeheartedly and unquestioningly commit yourself to some ideal, these questions are settled for you: you no longer have to be uncertain, confused, ambivalent, or adrift.[10]

Third, fanaticism can gratify certain epistemic and social anxieties. Again, philosophers and psychologists have argued that people crave (to different degrees) certain epistemic goods that fanaticism would provide. These include cognitive closure, relief from uncertainty, relief from ambiguity. And people also crave social goods that fanaticism can provide: a sense of belonging, community, shared purpose.[11]

But here's the problem: many of those things can also be done by non-fanatical forms of commitment, such as the form of commitment that I call *devotion*.

6. Devotion vs. fanaticism

What's devotion? Elsewhere (Katsafanas 2023a), I have defended this account of devotion:

---

[9] There is a vast literature on these points. Some useful introductory texts include Baumeister and Vohs (2002); Steger et. al. (2008); Wolf (2010); Kauppinen (2015); Taylor (1989); Metz (2013); Kruglanski et. al. (2014).
[10] There are some accounts of fanaticism that take this to be a primary motivation for fanatics. See, for example, Adorno et. al. (1950), Zmigrod et. al. (2020), and Kruglanski et. al. (2009)
[11] Again, there is a massive literature. See Jensen et. al. (2018), Hoffer (2010), Stern (2004), Stern and Berger (2015).

> **(Devotion)** An agent is devoted to an object (a person, state of affairs, value, goal, etc.) iff (i) she is committed to engaging in the relevant ways with that object, (ii) this commitment is dialectically invulnerable and (iii) she is disposed to maintain the dialectical invulnerability of this commitment.

A commitment or attitude is *dialectically invulnerable* when there is no argument, distinction, clarification, or other dialectical move that would dislodge it or cause it to dissipate. Showing the commitment or attitude to be irrational, inconsistent with other commitments, in conflict with moral principles, lacking adequate justification, and so forth has no effect.

So, on my view, devotion differs from ordinary commitment in an epistemic respect: it is dialectically invulnerable, and the person aims to maintain that invulnerability.

Note that, as I understand it, devotion doesn't have to involve commitments that override all other commitments. I can be devoted to philosophy but readily override this in order to respect commitments that I regard as more important, such as saving a life or helping a friend. Devotion differs from ordinary commitment in this epistemic respect, not in terms of strength or lexical priority.

Let's consider how devotion relates to fanaticism. Like the fanatic, the devoted agent manifests an unusually robust form of commitment to his object. Like the fanatic, the devoted agent's object will not be susceptible to being undermined by justificatory reasoning. And, like the fanatic, the devoted agent *might* be willing to impose costs on himself and others.

But this last feature will be accidental: I can be devoted to yoga or to a friendship without being willing to impose extreme costs on others so as to secure these goods. The degree of sacrifice that the object of devotion warrants will depend on the content as well as the lexical ranking of that object.

When we consider goals that have high lexical rank and whose content requires sacrifice (of oneself and others), devotion to these goals can look *very* similar to fanaticism. For example: many political and social activists—many abolitionists, for example—are devoted to their goals. In addition, many of them think that their goals are so important that they should outweigh other goods, warrant sacrifices, and so on. But the mere fact that a person is devoted to an overridingly important goal, and thus willing to undertake sacrifices and impose costs so as to sustain it, does not entail that they are fanatical.

With this distinction between fanaticism and devotion at hand, we can ask some questions. Does the fact that a person is willing to sacrifice himself and others for some value entail that he is fanatical? No, because acceptance of the value may legitimate the sacrifices. Does the fact that a person insulates his value from the effects of critical reflection entail that he is fanatical? No, because he may be devoted to the value rather than fanatical about it. Does the fact that a devoted person's value is reprehensible, misguided, or mistaken entail that he is fanatical? No, because you can be devoted to reprehensible ideals, and you can be fanatical about good ideals.

Thus, the distinction between devotion and fanaticism gives us a useful way of teasing apart importantly different considerations.

7. What's so bad about fanaticism?

I have been claiming that fanaticism is a way in which devotion can be corrupted. But I haven't yet explained what distinguishes ordinary devotion from fanaticism, making the former good and the latter bad.

Both can involve certain forms of closed-mindedness, commitment to sacred values, and shielding the object of commitment from the typical effects of justificatory reasoning. Both can involve extreme forms of sacrifice. Both can involve having values, commitments, or ideals that override all others.

But fanaticism involves a form of motivated reasoning. Familiarly, our theoretical reasoning about what to believe is often influenced by considerations that are unrelated to truth and falsity. Sometimes, anxiety, fear, desire, and so forth influence our reasoning, as when the self-deceived person can't fully consider evidence that might threaten his misguided beliefs. Just so, fanatical commitments are responsive to a particular kind of background motivation. This shows up not just in its origins of the fanatical commitment, but also in the ongoing ways that the person relates to his object of commitment.[12]

In Katsafanas 2023a, I have defended the following account of fanaticism. A person counts as fanatical to the extent that:

1. Sacred values: she adopts one or more sacred values.[13]

---

[12] Compare Goldsmith: "fanaticism, properly understood, means more than just devotion to a cause—more, indeed, even than extreme devotion to a cause. Rather, a fanatic assumes a unique way of being in the world, one that brings with it a host of characteristics that go beyond mere political engagement or even political extremism. Fanaticism, to truly merit the description, must possess some combination of messianism, an inappropriate relation to reason, an embrace of abstraction, a desire for novelty, the pursuit of perfection, an opposition to limits, an embrace of violence, absolute certitude, excessive passion, as well as some intellectual pretension." (Goldsmith 2022; 4-5). I think some of the features in Goldsmith's long list are contingent. But I do think he's right that fanaticism isn't just devotion.

[13] In Katsafanas 2023a, I define sacred values as follows: let V1 be a value or normative commitment. Then V1 counts as sacred iff it has the following three features:

1. **Inviolable**: if V2 is an ordinary value, then it is prohibited to sacrifice V1 for V2, regardless of the quantities of V1 and V2.

2. **Incontestable**: It is prohibited to *contemplate* trading or sacrificing V1 for most or all other values.

3. **Dialectically Invulnerable**: The agent insulates her commitment to V1 from the effects of justificatory reasoning. That is, while the agent may think about V1's justification, consider objections to V1, consider alternatives to V1, engage in thought experiments with respect to V1, and so on, the agent does not stake her commitment to V1 on the outcome of this justificatory reasoning. There is no dialectical move that would disrupt the agent's commitment to V1.

2. Fragility of the self: she needs to treat the value as sacred in order to preserve her identity.

3. Fragility of the value: she takes the value's status as sacred to be threatened when it is not widely accepted.

4. Group identity: she identifies herself with a group, where this group is defined by shared commitment to the sacred value.

Fanaticism's bad attributes come from features (2) and (3): fragility of self and of value. These forms of fragility are bad in two ways: first, they are bad in their effects—they dispose the agent toward intolerance. Second, they are bad in that they generate a defective form of identity (a defective relation to self).

To see this, let's first start with an explanation of what fragility of the self is. (2) claims that the person needs to treat a value as sacred in order to preserve her identity. In other words, the person believes that if they fail to sacralize the value (i.e., fail to treat it as inviolable, incontestable, and dialectically invulnerable) then their identity will be compromised or lost. In general, identities are in part constituted by deep commitments. So there's nothing wrong with seeing certain commitments as an essential, nonoptional component of who one is. For example, a devoted agent could see the devotion as integral to her identity: if she gave it up, or if it were threatened to a sufficient degree, she wouldn't be the same person. So there's a sense in which even the devoted agent needs to preserve her devotion in order to preserve her identity. Someone who devotes her life to environmental activism might not be able to understand herself as the same person independently of that end.

In short: there's nothing objectionable or problematic about seeing certain commitments as essential constituents of who one is. So what's the difference between merely treating something as an essential constituent of your identity and being fragile with respect to that constituent?

The key difference is the *unwarranted, exaggerated perception of threat*. Some of these threats come from disagreement or dissent. The devoted, non-fanatical agent need not see his value as threatened by dissent and disagreement. The fanatical agent does see his value as threatened by dissent. He demands universal acknowledgement of his value as sacred. Anything less is seen as destabilizing the value's status as sacred. In other words, the fanatical agent perceives mere disagreement as a threat. Rather than living with disagreement, living with the possibility that others will reject what he holds dear, and tolerating this dissent, he has a hostile, defensive response to this disagreement.[14]

And, relatedly, some of these threats come from contingency. The fanatic tends to be unable to acknowledge other ways of life, other deep commitments, as equally appropriate. The devoted environmentalist might admit a plurality of appropriate objects of devotion, not seeing this as a threat to his own devotion (after all, there are plenty of pressing social and political concerns other than environmental threats, and it can make sense to devote oneself to one of these instead of to

[14] For defense of this claim, see Katsafanas 2023a. In Chapter 7 of that work, I argue that the fanatic interprets non-acceptance of his sacred values as a threat to the standing of those values. For analogous claims, see Berger 2018, Doosje et al 2013, Jonas et al 2014, and Tietjen 2023b.

environmental causes). But the fanatical environmentalist won't be able to do this; he won't be able to acknowledge the contingency or optionality of his own commitment.

So: the fanatic feels the need to secure his value against threats. There's nothing wrong with securing one's value against real threats. That can be necessary. But the fanatic has an unwarranted, exaggerated sense of threat: he sees threats in the mere existence of disagreement and in the mere existence of contingency (the sense that there are multiple, incompatible, permissible objects of devotion).

Ruth Rebecca Tietjen (2023a) puts it this way: it's not really fragility as such, but *fear* of fragility, an attempt to eliminate threats to fragility, that's the problem. That is: if we think of fragility in a certain way, all of our values are fragile. We can give them up; we can fail to realize them; we can see them opposed or rejected or just ignored by others. This is a fact of life. Many of us are willing to accept this, seeing our values as fragile but not having recognition of this fragility motivate a hostile response. But the fanatic can't tolerate contingency, disagreement, dispute. The mere existence of disagreement and contingency is seen as threatening.

That's what differentiates a *fragile* identity and a nonfragile one. Both the fragile and nonfragile agent can see a core commitment as an essential component of their identity; both can attempt to secure that commitment against certain kinds of threats. But the fragile agent has an *exaggerated* sense of threat, seeing threats where there are none, and seeing mere disagreement as threat.

Empirical research supports this point: fanatics tend to display a heightened sense of threat, seeing themselves and their communities as facing "intense trouble, difficulty, or danger" (Jensen et al 2018: 7). Importantly, these individuals typically take mere disagreement or mere non-acceptance of their sacred values as constituting a threat.[15] For one example, consider Anders Breivik. Breivik carried out a bombing and mass shooting in Norway in 2011 and was widely characterized as a fanatic. Shortly before carrying out his attacks, which left 77 people dead, Breivik distributed a 1518-page manifesto that described the various ways in which he saw traditional nationalist values, which he regards as sacred, as threatened by Islam, political correctness, feminism, multiculturalism, and a host of other putative enemies. The examples of fanaticism offered at the beginning of this paper—the insurrectionist, the jihadist, the white nationalist—are analogous: there, too, we often find apocalyptic theories and a sense of widespread threat to their sacred values.

Other philosophers have made similar points. Josh DiPaolo claims that the fanatic treats disagreement (especially higher-order disagreement about what counts as evidence) "not as a reason to rethink their worldview but as a threat to their identity" (2020: 219). The fanatic feels the need to eliminate uncertainty and disagreement about his commitments. He is rigid with respect to them. He cannot see them as contingent, optional, or questionable. And Thomas Szanto (2022) also sees *threat perception* as an integral constituent of fanaticism. He writes, "So, what is the formal object of fanaticism? Or, what individuates fanaticism as the particular kind of affective phenomenon it is, distinguishing it, say, from hatred (an attitude that similarly aims at the elimination or social exclusion from of its targets…)?" (2022: 194). In other words, what makes the fanatic think that the target is worthy of fanatical reactions rather than just hatred or moral outrage? Szanto's answer is that the target *threatens:* "The formal object of fanaticism is the *threat* to the sacred value issuing from

[15] See Berger 2018, Doosje et al 2013, Jonas et al 2014, Stephan and Stephan 2000, Tietjen 2023b, and Katsafanas 2023a.

harm, questioning or dissent from the target group. It is not, as one might initially think, the alleged evilness of the target-group, or the hate-worthiness of their values or ideology" (2022: 194).

Fanatics are antagonistic. Sometimes people who are devoted are also antagonistic: they see that their value is threatened, and they respond. But antagonism isn't built into devotion. It is built into fanaticism.

With that in mind, let's now consider what might be problematic about fragility.

Let's start with the *effects* of fragility. Fragility transforms even innocuous commitments into problematic ones. Why is this? Because thinking of your value as fragile makes any disagreement or dissent look like a threat. So it involves an antagonistic, hostile stance toward others: their nonacceptance of the value is viewed as a threat to be dealt with. And the fragility of self makes any threat to the value look like a threat to the self. So mere disagreement is interpreted as hostile attack. That's what is bad about fanaticism: it takes innocuous disagreement (and innocuous contingency) and interprets it as threatening.

Second, fanaticism is an existential vice, a defect of identity. Notice that fanaticism is connected to identity. Being a fanatic about something involves making that thing integral to your identity, to your sense of who you are. The object of commitment becomes a non-contingent component of your identity. As I mentioned above, there's nothing wrong with that. It's fine to ground your identity in some kind of commitment to an ideal. Devoted agents do this. But, as I just explained, there's a kind of fragility, fear, or antagonism built into fanaticism. Again, any value will be open to threats. But the fanatic has a *heightened or unwarranted or exaggerated* sense of threat.

One good way to see this is to think of how fanatical *love* for a person differs from non-fanatical love. Imagine a person who is passionately, unwaveringly, resolutely committed to his lover. He worries about threats to her. He wants to preserve her well-being. He undertakes sacrifices for her. He prioritizes her over others. What would it take for this love to become not just devoted but fanatical? Not just that the devotion is part of his identity; we expect that, it's normal. Not just that the devotion is of central importance; again, we expect that. But suppose the person sees his beloved as needing to be acknowledged as superior by everyone. Imagine he feels threatened, feels the status of his beloved as diminished, when others do not see her as preeminent. This begins to look more pathological. By itself, it doesn't constitute fanaticism (because we don't yet have the intolerant attempt to eliminate the dissent). But it starts to get closer.

8. Fanaticism as a virtue or a vice?

I've been arguing that fanaticism is bad both in terms of its effects (hostile intolerance) and in terms of the type of relation to self and value that it involves. Let's now connect this to Battaly's argument that fanaticism is normatively neutral.

Battaly distinguishes two types of virtues (and hence two things we could mean by asking whether fanaticism can be a virtue):

- Effects-virtue: "a stable disposition will be an effects virtue if and only if it produces a preponderance of good effects" (Battaly 2023: 17). This can be contextualized: what counts as an effects-virtue in one context may not count as an effects-virtue in sufficiently different contexts (Battaly 2023: 17-18).
- Liberatory virtue: "for a disposition to be a liberatory virtue, it must have liberatory content: the agent must be aware of oppression, and must be motivated to resist it. They must also be motivated to make progress toward liberation and the future flourishing of all." (Battaly 2023: 19) [16]

Battaly wants to say that fanaticism can be an effects-virtue in certain contexts (2023: 17-18). In other words, there are certain circumstances in which fanaticism is a stable disposition producing a preponderance of good effects. For example, in a context in which slavery is legal but potentially capable of being rendered illegal by devoted action, fanatical abolitionism can produce a preponderance of good effects and hence count as an effects-virtue.

This seems right to me; I have no qualms with the idea that fanaticism, as a form of devoted action, can produce good effects. But notice that devotion without fanaticism can do the same thing, and won't involve the problems. So saying that fanaticism can be an effects-virtue sounds to me like saying that paranoia can be an effects-virtue. Sure, there are some circumstances in which being paranoid is going to reliably produce good effects. But it's still a vice in a broader sense.

Of course, it's open to Battaly to respond by claiming that devoted abolitionism is not sufficient for producing a preponderance of good effects, but that fanatical abolitionism is. In other words, Battaly could argue that fanaticism somehow achieves effects that devotion alone cannot. But this would require a substantial argument; on the face of things, it's not obvious why devotion couldn't produce the same effects as fanaticism.

Let's move on to liberatory virtue. Battaly argues that fanaticism can be a liberatory virtue. To return to her example, certain abolitionists fulfill this condition. They have an awareness of oppression and a motivation to resist it, but also fulfill the conditions for fanaticism. But again, even if this were true devotion would be better: it would get you the awareness of oppression and the motivation to resist it, but wouldn't involve the problems.

And there's a further, deeper problem: notice how precarious fanatical liberatory tendencies would be. If we're talking about a genuine case of fanaticism, it will involve an exaggerated sense of threat and thus can easily swing into a form of oppression on its own. We can all think of examples of social and political movements that start out looking liberatory, but deform into oppression—the classic example is the French Revolution, which devolves into the Terror, but that is just one instance among countless others. So fanatical devotion to liberatory ends is perhaps better than indifference to them, and is definitely better than opposition to them. But it is still a vice, disposing the agent to manifest intolerance and an existential defect.

9. Conclusion

[16] She also lists a third type of virtue, which she defines as follows: for a stable disposition to be a *character virtue*, it must be "partly constituted by good motives and good values"—it's not enough for it just to produce good effects (Battaly 2023: 18). She doesn't take fanaticism to be a character virtue, so I do not discuss this type of virtue above.

So, what's so bad about fanaticism? Is it bad at all? Cassam claims that fanaticism is bad because it involves commitment to a morally reprehensible end. Battaly claims that fanaticism is neutral: it can be bad, but if it is this will be due to the badness of the fanatic's end rather than to the fact that the person is fanatical. I've argued that these accounts mislocate the concern: we should focus not on the fanatic's object (their end, goal, ideal, etc.), but on the fanatic's psychological profile. Battaly is right that the fanatic can adopt good or neutral objects, but nonetheless I've argued that fanaticism involves an existential defect and a disposition toward violent intolerance. In that sense, fanaticism is bad. Nonetheless, we can be nuanced in our judgments about particular cases of fanaticism. Fanatics can be directed toward morally praiseworthy goals, and the good effects of fanaticism can in some circumstances outweigh the bad aspects.

In making this argument, I've distinguished between devotion and fanaticism. I see this as a significant advantage of my account. It gets us what we want: the idea that there's an extreme form of commitment that is praiseworthy and that is often necessary to achieve important social and political goals. But my account preserves the idea that fanaticism is a vice. And it allows us to say that there can be cases of *devotion gone bad,* devotion-turned-badly-fanatical.

I will close by mentioning one further point. Aside from offering us the resources to distinguish devotion from fanaticism, my account has a further advantage: it gives us a way of addressing fanaticism. If the core problem with fanaticism is just that the fanatic is directed toward objectionable ends (Cassam's view), then the solution is to change people's ends. But if the core problem with fanaticism is an existential one—a defect of character, of one's way of relating to value quite generally—then the solution is to change people's characters. That is no easy task—but, if we have any hope of mitigating the strife and discord caused by fanaticism, it is a crucial one.[17]

[17] For helpful comments on this paper, thanks to two anonymous referees as well as audiences at UConn and Boston University.

References

Adorno, T.W., Else Frenkel-Brunswik, Daniel J. Levinson, and R. Nevitt Sanford (1950), *The Authoritarian Personality*. New York: Harper and Row.

Battaly, Heather (2023), "Can Fanaticism be a Liberatory Virtue?," *Synthese* 201(6): 1-27.

Baumeister, R.F. and Vohs, K.D. (2002), "The Pursuit of Meaningfulness in Life," in C.R. Snyder and S. J. Lopez (eds), *Handbook of Positive Psychology*. Oxford: Oxford University Press, 608-618.

Berger, J.M. (2018), *Extremism*. Cambridge: MIT Press.

Cassam, Quassim (2022a), "Fanaticism: For and Against," in *The Philosophy of Fanaticism: Epistemic, Affective, and Political Dimensions,* Routledge 2022.

Cassam, Quassim (2022b), *Extremism: A Philosophical Analysis*. New York: Routledge.

Colas, Dominique (1997), *Civil Society and Fanaticism: Conjoined Histories,* trans. Amy Jacobs. Stanford: Stanford University Press.

Crosson, Frederick J. (2003), "Fanaticism, Politics, and Religion," *Philosophy Today* 47(4): 441–7.

Deleyre, Alexandre (2009), *The Encyclopedia of Diderot & d'Alembert Collaborative Translation Project,* trans. by Stephen J. Gendzier. Ann Arbor: Michigan Publishing. <http://hdl.handle.net/2027/spo.did2222.0001.305>. Trans. of *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, vol. 6. Paris, 1756.

DiPaolo, Joshua (2020), "The Fragile Epistemology of Fanaticism," in Michael Klenk (ed.), *Higher-Order Evidence and Moral Epistemology*. New York: Routledge, 217-235.

Doosje Bertjan, Annemarie Loseman, and Kees van den Bos (2013), "Determinants of Radicalization of Islamic Youth in the Netherlands: Personal Uncertainty, Perceived injustice, and Perceived group threat," *Journal of Social Issues* 69: 586–604.

Goldsmith, Zachary (2022), *Fanaticism: A Political Philosophical History*. University of Pennsylvania Press.

Griffen, Charles (2009), "John Brown's 'Madness'," *Rhetoric and Public Affairs* 12(3): 369-388.

Hoffer, Eric (2010), *The True Believer: Thoughts on the Nature of Mass Movements*. New York: Harper.

Jensen, Michael A., Anita Atwell Seate, and Patrick A. James (2018), "Radicalization to Violence: A Pathway Approach to Studying Extremism," *Terrorism and Political Violence,* DOI: 10.1080/09546553.2018.1442330.

Eva Jonas, Ian McGregor, Johannes Klackl, Dmitri Agroskin, Immo Fritsche, Colin Holbrook, Kyle Nash, Travis Proulx, and Markus Quirin (2014), "Threat and Defense: From Anxiety to Approach" in James Olson and Mark Zanna (eds.), *Advances in Experimental Social Psychology,* vol. 49. San Diego: Elsevier Academic Press, 219–286.

Katsafanas, Paul (2019), "Fanaticism and Sacred Values," *Philosophers' Imprint* 19(17): 1–20.

Katsafanas, Paul (2023a), *Philosophy of Devotion: The Longing for Invulnerable Ideals*. Oxford: Oxford University Press.

Katsafanas, Paul (ed.) (2023b), *Fanaticism and the History of Philosophy*. New York: Routledge.

Katsafanas, Paul (2023c), "An Introduction to the Philosophy of Fanaticism," in Katsafanas (2023b), 1-18.

Kauppinen, A. (2015), "Meaningfulness," in G. Fletcher (Ed.), *Routledge handbook of the philosophy of well-being*. New York: Routledge, pp. 281-291.

Klein, Lawrence E. and Anthony J. La Vopa (1998), *Enthusiasm and Enlightenment in Europe 1650–1850*. San Marino, CA: Huntington Library.

Kruglanski, A. W., & Fishman, S. (2009). "The need for cognitive closure," in M. R. Leary & R. H. Hoyle (eds.), *Handbook of individual differences in social behavior*. The Guilford Press, 343–353.

Kruglanski, A.W., Michele J. Gelfand, Jocelyn J. Bélanger, Anna Sheveland, Malkanthi Hetiarachchi, and Rohan Gunaratna (2014), "The Psychology of Radicalization and Deradicalization: How Significance Quest Impacts Violent Extremism," *Political Psychology* 35: 69-93.

La Vopa, Anthony J. (1997), "The Philosopher and the *Schwärmer:* On the Career of a German Epithet from Luther to Kant," *Huntington Library Quarterly* 60: 85–115.

Martinich, A.P. (2000), "Religion, Fanaticism, and Liberalism," *Pacific Philosophical Quarterly* 81: 409–25.

Metz, Thaddeus (2013), *Meaning in Life.* Oxford: Oxford University Press.

Oates, Stephen B. (1984), *To Purge This Land With Blood: A Biography of John Brown.* University of Massachusetts Press.

Olson, Joel (2007), "The Freshness of Fanaticism. The Abolitionist Defense of Zealotry," *Perspectives on Politics* 5 (4), 685–701.

Passmore, John Arthur (2003), "Fanaticism, Toleration, and Philosophy," *Journal of Political Philosophy* 11(2): 211–22.

Reynolds, David S. (2006), *John Brown, Abolitionist.* New York: Vintage.

Rousseau, Jean Jacques (1979), *Emile,* trans. Allan Bloom. New York: Basic Books.

Szanto, Thomas (2022), "Sacralizing Hostility: Fanaticism as a Group-Based Affective Mechanism," in Towsend et al (2022), 184-212.

Steger, Michael et. al. (2008), "Understanding the Search for Meaning in Life: Personality, Cognitive Style, and the Dynamic Interplay Between Seeking and Experiencing Meaning," *Journal of Personality* 76(2): 199-228.

Steinberg, Justin (2023), "Superstition, Enthusiasm, and Radical Enlightenment from Hobbes to Hume," in Katsafanas (2023b), 67-82.

Stephan, Walter G. and Cookie White Stephan (2000), "An Integrated Threat Theory of Prejudice," in Stuart Oskamp (ed.), *Reducing Prejudice and Discrimination.* Hillsdale, NJ: Lawrence Erlbaum, 23–46.

Stern, Jessica (2004), *Terror in the Name of God: Why Religious Militants Kill.* New York: Harper.

Stern, Jessica and John M. Berger (2015), *ISIS: The State of Terror.* New York: Harper Collins.

Taylor, Maxwell (1991). *The Fanatics: A Behavioural Approach to Political Violence.* Oxford, Brassey's.

Taylor, Richard (1999), "The Meaning of Life," *Philosophy Now* 24:13-14.

Tietjen, R. R. (2021), "Religious zeal as an affective phenomenon," *Phenomenology and the Cognitive Sciences,* 20(1), 75–91.

Tietjen, Ruth Rebecca (2023a), "Religious Zeal, Affective Fragility, and the Tragedy of Human Existence," *Humanistic Studies* 46: 1-19.

Tietjen Ruth Rebecca (2023b), "Fear, Fanaticism, and Fragile Identities," *Journal of Ethics* 27(2): 211-230.

Toscano, Alberto (2010), *Fanaticism: On the Uses of an Idea.* New York: Verso.

Townsend, Leo, Ruth Rebecca Tietjen, Hans Bernhard Schmid, and Michael Staudigl (eds.), (2022), *The Philosophy of Fanaticism: Epistemic, Affective, and Political Dimensions.* New York: Routledge.

Voltaire (1984). *Philosophical Dictionary.* New York: Penguin.

Wolf, Susan (2010), *Meaning in Life and Why it Matters.* Princeton: Princeton University Press.

Zmigrod, L., Rentfrow, P.J., & Robbins, T.W. (2020), "The partisan mind: Is extreme political partisanship related to cognitive inflexibility?," *Journal of Experimental Psychology: General* 149(3): 407-418.