LUND
UNIVERSITY

If One Thinks About Raping Someone in the Woods and No One Knows, is it Wrong?

An Argument for the Moral Relevance of Thoughts*

Om en tänker på att våldta någon i skogen och ingen vet, är det fel?
Ett argument för tankars* moraliska relevans

Neo Wikman

Department of Philosophy

Bachelor's Thesis

FPRK01:2

Spring term 2026

Supervisor: Jakob Werkmäster

Abstract

There are arguments around the nature and boundaries of the moral domain, with this current paper aiming to delineate one of these arguments. With a basis in George Sher's (2021) discussion of cognitions (*beliefs*, *fantasies*, *attitudes* and *some thoughts*), the paper defends the notion that they are, in fact, morally relevant. By combining the logic of truth-aptness with the phenomenology of agency, the paper demonstrates that private mental states can be regarded as genuine acts. These acts are shown to be morally relevant due to them being (i) directed, (ii) allowing for psychological agency, and (iii) endorsing a moral object. From this, cognitions can constitute and produce genuine, unfelt harms towards the subject and frustrate desires. It would indicate that the privacy of the mind provides no exemption from moral accountability; to think is to act, and to act is to be subject to the moral domain.[1]

Keywords: *Doxastic wronging, unfelt harm, doxastic voluntarism, cognitions, fantasy*.

[1] I want to express gratitude to my supervisor, Jakob Werkmäster, for his guidance and presence during this project. I also want to acknowledge numerous acquaintances for their support by discussing this topic and providing their feedback. A thank you to a cursory cirrocumulus is in order as well, for passing by and raining insight.

1. Introduction

Debate over the boundaries of the moral domain is common, with there being no clear, uncontroversial conclusion regarding the distinction between the public and private spheres. Two main perspectives emerge: one views the mind as a private domain beyond moral scrutiny, while the other finds moral relevance in private mental states. Arguments and cases exist on both sides. For example, agents are said to lack volition over their thoughts, so they cannot be held responsible for them. On the other hand, however, individuals may feel wronged by someone's beliefs about them; if such beliefs are wrong, then private mental states are morally relevant. Yet, can thoughts and beliefs be morally evaluated? Are simply holding misanthropic beliefs like sexism or racism morally wrong, or must action follow?

It may seem odd to view thoughts as wrong. If it is, consider this case:

> *A Thought in the Woods* (ATW): Rasmus is out hiking alone in the woods. While hiking, Rasmus gets a thought in his head that says, "I want to rape someone."[2]

The thought[3] that Rasmus has is not acted upon. Is the thought as described in ATW morally relevant and thus condemnable? This paper seeks to answer that question by clarifying both the role and the relevance that thoughts have in the moral realm. The main thesis is that thoughts can, in certain cases, be morally relevant—a position supported by the concept of doxastic wronging.[4]

A case for why thinking is morally relevant is delineated in Section 2. This is established through arguing for doxastic voluntarism, the idea that beliefs are subject to voluntary control (Boespflug & Jackson, 2024). I argue that there is a case for an agent to exert some control over

[2] There is no specified target of the thought, but it is still directed (see Section 2). A target can exist that the propositional content of the thought (i.e. the meaning of the content and its truth-aptness) is referring to, while the subject referred to does not need to be specified. Other thoughts, such as misogyny, also have a nebulous target. Misogynistic thoughts are still directed, in this case, towards women, but the thoughts may not be specifically targeted. The thought could still be considered to have an identifiable target, just not a specific one. Targets of the thought can still be wronged by the thought causing harm, as the subject can be considered to undergo a setback to one of their fundamental interests (Zhou, 2022). Exempli gratia, the statement "Women should be less valued than men" would be aberrant if it were assumed to be bound by temporality; it is reasonable to hold that the statement would include people who are yet to be born. But one cannot hold that these people are identifiable. This would show that there may not be a need for either a specific or identifiable target for a thought to harm. From this, it will be taken that unspecified thoughts, such as "I want to rape someone", have a target, not just a specific one. What wronging and harmful thoughts entail will be discussed in Section 3.

[3] See Section 2.

[4] The premise for the argument is intended as *a priori*. There is a consequentialist argument to be made that those who engage with wrongful thoughts are more likely to engage in actual acts that correspond with the thoughts, e.g. rape or bigotry. Yet, this is not what is in focus here. The aim is to ascertain whether thoughts *qua* thoughts can be wrongful.

their beliefs. Thoughts that are spontaneous can likewise qualify for moral evaluation. The paper establishes that the necessary and sufficient conditions for a thought to be morally relevant are *content*, *endorsement*, and *psychological agency*. In Section 3, a discussion on whether the effect thinking will be said to have is in fact truly relevant through wronging. By establishing an account for unfelt harm, cognitions that are *directed* can be understood as wrong.

2. Thoughts* as morally relevant

It would be absurd to hold that all thoughts are morally relevant, since some thoughts do not encroach on the moral realm, due to their content being amoral. A comparison could be made between values. Not all values are moral; there is a difference between moral and aesthetic values. Thoughts that do not pertain to the moral realm are not pertinent to this discussion. Aesthetic observations or pondering breakfast alternatives would not be seen in light of a moral nature. It would equally be reasonable to hold that "there are a dozen sheep on the lawn" would be just as amoral if it were stated or if it were thought; their content is not of a moral nature. It appears that, of necessity, the content of a thought is pertinent to whether it can be morally evaluated.

Still, there is no clean dichotomy between morality and amorality. Statements such as "Charles Manson was a great songwriter" or "*Salò, or the 120 Days of Sodom* is a terrific film" have the characteristic of pure aesthetic content, yet they are still clearly not amoral propositions.

For this discussion, it will be held that "moral" refers to objects that relate to codes of conduct that are endorsed by someone, no matter whether they are prohibited or permitted (Gert & Gert, 2025). This will be understood as objects that produce harm or benefit. In Section 3.1, this will be elaborated on further, but for the moment, this definition will suffice.

Before proceeding, it is important to clarify the terms used. *Thinking* refers to mental activity (Antony, 2009), which does not require physical processes and involves deliberate engagement in analysis, consideration, or contemplation. *Thoughts* are the cognitive products of thinking but may also arise from external stimuli or internal associations. A thought need not be accepted as true; for example, one could imagine a pink talking turtle without believing it to exist. Thoughts are typically ephemeral and can be irrational. *Beliefs* are mental attitudes that represent a commitment to the truth of a proposition (Heiphetz et al., 2021). They serve as dispositional states, persisting over time and influencing interactions. Beliefs contain distinct

cognitive attitudes (Heiphetz et al., 2021), and these form a subset of propositional attitudes of how the world is and might be (Shah & Velleman, 2005).

To evaluate the mental realm with precision, the paper adopts *cognitions* as a functional hypernym encompassing *beliefs*, *fantasies*, *attitudes,* and *some thoughts*. While ordinary discourse often utilises the term' *thoughts'*, this construct[5] is problematic because it conflates unbidden, passive mental events with directed, intentional cognitive processes. Additionally, relying on *doxastic*, which only refers to beliefs (Basu & Schroeder, 2018), imposes a specificity that can obscure the moral weight of non-belief states, such as the fantasy in ATW. Unlike thoughts, the construct of cognitions allows for a distinction between involuntary associations and mental acts that exhibit differing degrees of *psychological agency* (Steup, 2012). To establish moral responsibility, a distinction must be drawn between passive associations and mental acts that involve psychological agency. Cognitions allow agents to exercise control over them (see Section 2.1) and are apt to have morally relevant content, due to their having the potential of endorsing an object (see Section 2.3).

Still, it is reasonable to object to this. The term "doxastic" refers to propositions that commit to truth (Singh, 2026), whereas other cognitions may lack this truth-commitment. As mentioned, agents can entertain a cognition without asserting it to be true. So if cognitions are not truth-apt, can they even be evaluated morally or be likened to the doxastic? Though, as Jamin Asay (2021) argues, cognitions need not express an explicit truth-claim for them to be truth-apt, since “when we use propositions, we (attempt to) do things with the truth” (p. 12). The content of the propositions then contains the concept of truth, and to possess it is simply to possess the cognitive equipment necessary to entertain propositions. As this shows, the propositions involved in cognitions, even though they do not make explicit truth-claims such as beliefs, are in fact truth-apt. For every propositional cognition, truth is a necessary component; the cognition itself proposes truth conditions.

Withal, the normative weight of a cognition does not reside solely in its truth-claim; it is affected by its directed representational state. The cognition in ATW may not contain explicit truth assertions, but its inherent structure is built upon its truth-aptness and the proposition itself being necessary for moral evaluation. This would appear to quell objections to the paralleling

[5] A *construct* refers to an ideal object, i.e. an object of the mind (Willer & Webster, 1970). These would typically be abstract objects such as “intelligence” or the symbol of “2”. It is then conceptual entities which are not directly observable but can be inferred through observation. E.g. the centre of mass of an object does “exist”, but not to the same extent as the object itself does.

of cognitions to the doxastic in the sense of allowing them to be susceptible to the framework of doxastic wrongings. This will be further discussed in Section 2.4.

From what has been established thus far, the paper proposes the following thesis:

> Th: Cognitions are morally relevant iff they voluntarily endorse moral objects.

If Th is true, then it would show that cognitions, such as beliefs and certain thoughts, can be morally relevant. Th and its relevance will be elaborated upon in Section 2.1–2.4.

Cognitions can be exhibited publicly, but can also be purely private. Private cognitions lack an external aspect, such as expression or manifestation in action. Public cognitions, which are expressed verbally or in actions, will not be discussed further in this paper.[6] This does not mean that cognitions are not actions. One *can* exercise agency over one's cognitions, such as mulling over a name one does not remember or deciding to stop and think about something else.

Some cognitions are both private and public—they are composite occurrences. ATW discusses a purely private cognition. If acted out, it would become a composite and not a purely private occurrence. Thus, an amendment to the Th seems appropriate. In this paper, "cognitions" will refer to purely private cognitions unless otherwise specified.

ATW is a purely private cognition, but is that morally relevant? Answers vary, but one person's thoughts on the matter are worth mentioning.

2.1 A Case for Doxastic Voluntarism

The fundamental challenge to the moral relevance of the mental is the problem of control. George Sher (2021) contends that the mind is "best regarded as a morality-free zone. Within that realm, no thoughts or attitudes are either forbidden or required" (p. 1). His thesis is as follows:

> Wild West of the Mind: For any cognition C[7], in an agent's mind, C is not morally wrong.[8]

[6] It is superfluous to argue for public cognitions. If a private cognition could be understood as wrongful, then it would follow that a public cognition also has the ability to be it.

[7] When discussing cognitions, Sher refers only to purely private ones. He does not stipulate what cognitions are, but it appears to encompass beliefs, attitudes, and fantasies (pp. 2–3), which is analogous to the definition of cognitions presented in this paper.

[8] There are a few caveats that could alleviate questions surrounding the Wild West of the Mind. Primarily, Sher (2021) holds that cognitions can be epistemically unjustified. Holding grudges against others purely on inference can be wrong. Secondly, cognitions have a bearing on one's character; it is not nice to find pleasure in another's pain (p. 5). His point is that these reasons, epistemic and aretaic, are the only forms of condemnation that are permissible. There are no moral reasons not to be vicious or irrational.

Even though Sher agrees with the idea that one can have agency over related mental activities (2021, p. 12), he holds that there is no case for doxastic voluntarism. The notion that one can have direct control[9] of one's beliefs is "implausible" according to him (p. 89). Belief is, as Sher argues, generally determined by a person's reason (p. 90). Beliefs can be irrational, but are usually construed on evidence: it is wet on the ground, so it must have rained. This is an automatic transition from evidence to belief, where an agent understands an evidential connection between objects. Beliefs are thus determined by the available evidence for an agent. These beliefs can be subject to irrational influences, yet their existence is evidence for them (p. 90). Thus, it can be held that "no one [can] ever acquire beliefs at will" (Alston, 1988, p. 268). One cannot will to believe '*p*' intentionally: you cannot believe that you have the ability to fly or that the United States is still a colony of Great Britain, just by deciding to do so. Alston (1988) argues that this is an effect of one lacking control over their *propositional attitudes*. Given the evidence, it is evidently true that you cannot fly or that the United States is not a colony of Great Britain. Thus, one does not have a doxastic choice. This suggests involuntarism.

Pamela Hieronymi (2006) proposes a similar conclusion, which, however, includes some control. She suggests that believing is to be understood as two separate activities, with an agent having varying degrees of control over each. First, is the activity of asking a question, by initiating an inquiry. Second, there is the evidence that is gathered, or the *constitutive reasons*, that settle a particular belief an agent can hold. One can have control over one's beliefs through direct control over the first, but for the second, it is only a matter of evidence of what belief is formed. Hieronymi states that there is control one can exercise, but one gains a certain understanding due to the available evidence. Consider the belief that it takes 30 minutes to get to your friend's house. If you pick up dinner beforehand, you would have to account for the time it takes to get there, which would subsequently alter the belief. By inquiring into the duration of the travel, the constitutive reasons enforce a certain, settled belief; it cannot take the same time to get to them by having to do some errands as without them.

If it is to be argued that cognitions can be wrongful, then having managerial control over them is not sufficient, since the constitutive reasons would be outside of volitional control. Actions are paradigmatically within one's control, while cognitions are construed not to be. In

[9] The exact extent of control "voluntary control" would entail is controversial. It is often understood as direct control, but could also figure as indirect control. Sher's (2021) analysis assumes that it is direct, and it will thence be assumed to be thus in this paper.

ATW, it would resultantly be understood that the cognition is not within Rasmus' control and would therefore not be considered relevant.

Though this position that beliefs are automatic is not as solid as it appears. Sher's view assumes that doxastic voluntarism would require a responsiveness to practical reasons, e.g. incentives. But a sufficient criterion of control can be understood through the capacity to be moved by an appreciation of reasons (Shah, 2002). The two-stage view is adapted for cases with a need for non-specific closure or the need to avoid non-specific closure (Kruglanski, 2004). This is a result of all the practical considerations that matter for whether to form a belief at all; the reasons do not matter for which belief one acquires. But an agent may hold epistemic motivation to acquire a specific closure and will not be satisfied with any answer (p. 6). It has a *directionally biased* influence on the epistemic process, where an agent has preferences for a particular answer to a question and thus would show epistemic agency. There are other reasons for forming beliefs that are not constitutive, with the foremost being one's own psychology.

Steup (2012) suggests that one *can* acquire beliefs intentionally through consciously believing certain propositions. If something is done intentionally without any prior intentions, such as shifting gears, then it is *implicit intentionality*. Steup suggests that beliefs may be an example of this as well. By being responsive to epistemic reasons–the evidence–then one can exercise control over one's beliefs (Steup, 2012, p. 156). Thus, it appears that agents can exert some control and regulate their beliefs.

The position of beliefs being automatic and directly causal in a predictable manner due to the evidence is also critiqued by Basu and Schroeder (2018). The portrayal of cognitions as being at the mercy of evidence is likened to a tornado being at the mercy of the atmosphere's instability. It is, in this view, just as odd to say that an agent ought to believe *'p'* as it is to say that a tornado should take another path (p. 8). Although the same evidence does not per se need to provide the same belief. One can be licensed to believe, upon seeing a shape in the sky, different propositions: (a) there are things in the air above, (b) there are birds flying ahead, (c) there is a handful of jackdaws flying above (Nelson, 2010, p. 87). The beliefs (a), (b), and (c) are all reasonable, though they are all subject to *background information*, *particular interests* and *needs,* which can influence an agent's settled belief. Even from this evidence, it does not follow that it needs to produce a belief. If you need to buy milk, then you do not need to acquire (a)–(c). One could consequently gain and hold beliefs that are unsubstantiated.

Another factor that matters for epistemic agency, barring evidence, could be moral reasons (Basu & Schroeder, 2018, p. 14). When answering a question about the proposition of a belief, the response comes down to a matter of one's own psychology. A settled belief is intimately

connected to one's epistemic motivations, the goals that one possesses with respect to knowledge (Kruglanski, 2004, pp. 30–31). Non-evidential considerations are influential, implying that there is no problem of control for cognitions.

On the question about making oneself believe that one can fly or rather *voluntary control*, Steup (2012) means that involuntarists conflate *volitional control*, the ability to make a decision, and *executional control*, the ability to carry out that decision (p. 146). For an involuntarist, it is assumed to be an *executional failure*, but Steup argues that it is actually a *volitional failure*. One cannot believe in having wings simply because one cannot *decide* to believe such a thing.

If the cognition Rasmus had in ATW did have an epistemic motivation that aligned with the cognition, then it would be reasonable to hold it to be of moral relevance. If the cognition had just popped into Rasmus' head and then he would have quickly dismissed it, then the scenario would be irrelevant. Yet, if the agent indulges in the cognition, then it is of moral relevance. The agent does not need to have an intention of having a rape fantasy, but by having it, they are implicitly manifesting intentionality through the engagement. This functions as an *intention in action*, as opposed to *prior intention*. Searle (1980) explains that "all intentional actions have intentions in action, but not all intentional actions have prior intentions" (p. 52). Cognitions do not need to have any prior intention, but they are still not automatic. The intentionality is embedded in the act itself, rather than being the execution of a mediated plan, similar to when one shifts gears or unscrews a toothpaste cap (Steup, 2012). However, if there was a lack of volition, then it would not be of moral relevance. Though, for the argument's sake, it will be taken that Rasmus *has* control over the cognition.

Still, Sher's view on morality does not presuppose any control; the notion of 'ought implies can' that can be utilised to argue that one is not responsible for acts that are outside of one's control, is not applicable to Sher. One can be held responsible for things outside of voluntary control, for example, negligent omissions or forgetting appointments (Sher, 2021, p. 11).

Following this, one can have at least some control over some cognitions, while spontaneous cognitions also exist. Still, even if some cognitions came about more deliberately than others, it is not detrimental for a cognition to be spontaneous. This will be expanded upon later in Section 2.4.

2.2 Dye-in-the-Water

If some cognitions are morally relevant and some are morally forbidden, then Sher (2021) means that we would be under a derivative obligation to perform acts we can control to avoid

the forbidden cognition (p. 92). Here, Sher introduces the argument of *Dye-in-the-Water*. It posits that there is a difficulty in establishing a discrete demarcation for cognitions, where "a prohibition against any single [cognition] is bound to diffuse itself among other [cognitions]" (p. 94). The inherent nature of cognitions and their permeability would subsequently mean that one is to avoid cognitions '*n*', 'o', and '*p*' that lead to the forbidden cognition '*q*'. If having cognition '*p*', which would lead to the murderous cognition '*q*', is not within voluntary control, then the cognitions are not morally relevant. Prohibitions for one cognition would lead to prohibitions against others.

Yet, as Samuel Director (2022) discusses, actions would face the same difficulty of demarcation. They lack as well as discrete boundaries. To kill your neighbour would mean that it is wrong for one to do things outside of the killing, such as acts leading up to the killing, e.g. buying a weapon, researching methods of disposing of a corpse, etc. (p. 425). These so-called preparatory acts would also be wrong.

Sher could mean that this is not the case for actions. He means that actions are generally more within one's control than cognitions. If φ-ing is said to be wrong, then the act preceding φ would also be considered to be wrongful. Though, where is one to draw the limit? It could be argued that only the acts that most likely lead to φ are wrong, and not the ones leading up to φ. This is because there may not be any causal connection between the acts. Buying a weapon would not be causal with killing one's neighbour. One can exercise control over one's actions; one can choose to kill one's neighbour, but one cannot be blamed for falling upon the cognition of killing them.

Still, permeability seems to be applicable to actions. If buying a weapon would knowingly increase the likelihood of killing one's neighbour, would it not be wrong to purchase it? Thus, if one tries to levy this argument against cognition, then it seems to be fair to use it against actions as well. Still, one can also bite the bullet and hold that if the cognition '*q*' is wrong, then all cognitions that knowingly lead to '*q*' are somewhat wrong (Director, 2022, p. 426). Further, all acts could be argued to stem from the mental, so the permeability of cognitions would have to be applicable to them as well.

If one takes a step back from this discussion, then a question may emerge. Why is there a need for demarcation to be available? For Sher, it is obvious; if one cannot control cognitions, then one cannot be accountable for reaching a specific one. If there is a difficulty in establishing a clear demarcation between a completed act (such as killing your neighbour) and a preparatory act (buying a weapon), how could it be established for cognitions? This will be discussed in Section 2.4.

The cognition in ATW may be spontaneous. Rasmus may be reminiscing about his old friends, what they did together, remembering that they went to the cinema a bunch. One of the films had a rape scene, and while reminiscing, Rasmus imagines himself as the perpetrator.

If Sher's argumentation is convincing, then all the cognitions preceding (friends, activities with friends, cinema and films) would then also not be permissible, which would be absurd. But, if Director is correct, then actions would face the same difficulty. By going outside, one can meet the neighbour and then kill them, or by going to the store, one can purchase a weapon to kill them with; is either of these acts, going outside or going to the store, wrongful? The paternalism that Sher is against would be pertinent for acts, as well as for cognitions.

Still, the focus is not on what cognitions occurred prior to or subsequent to having the cognition. It is cognition *qua* cognition that is of importance. The same would be true if one were to judge the killing of one's neighbour. It would be wrongful to do the acts preceding it, but it is the act *qua* act that one judges. As Section 2.1 exhibited, an agent can exert control over cognitions, as one can with actions. Thus, it would follow that cognitions applicable to morally relevant actions are.

From this, one can *prima facie* hold that the spontaneity of cognitions has no significant effect on their normative status.

2.3 What is Special About our Minds?

All this discussion has had an implicit implication about cognitions, or rather, where cognitions take place. In Director's (2022) paper, he argues against Sher's (2019) thesis[10] by asking a rather simple question, "What is special about our minds?" (Director, 2022, p. 427). The premise for Director is that it would not matter whether something is occurring in our minds or not. To show why this could be weird, consider the following case:

> *A Statement in the Woods* (ASW): Rasmus is out hiking alone in the woods. While out there, Rasmus publicly states, "I want to rape someone."[11]

ASW is contrasted with ATW, where the only difference between the cases is that in one of them the cognition becomes public, and the other remains private. It is stipulated that in both cases nobody will find out about what has occurred. If Sher were correct, then there would be a moral difference between the cases, something that Director finds to be untenable. In his work, Sher discusses the capacity to harm, where he could hold that the utterance has the capacity but

[10] Sher (2021) expands on Sher (2019). Both of their theses are the same.
[11] Both ATW and ASW are adapted from examples from Director (2022, p. 427).

not the cognition. Yet it is stipulated that nobody will hear Rasmus. And even so, it is possible to imagine a version where someone reads Rasmus's mind and is harmed by the cognition. Thereby, there appears to be nothing that demonstrates ATW, but not ASW, to be permissible. That a cognition is purely mental does not, in itself, seem to constitute a moral difference, despite Sher (2021) positing otherwise. And still, there are similarities between the cases. The most outstanding of which is that both are *endorsements* of a morally wrongful idea and action. Director (2022) posits that "it is *prima facie* wrong to endorse morally wrong ideas" (p. 429).

To understand this better, take the following case:

> *A Walk in the Woods*: Peter is out hiking alone in the woods. While out there, Peter thinks about hugging his cat.

Rasmus is identical to Peter in all aspects; they are generally nice to others and do similar activities. Yet, while Peter holds non-objectionable beliefs, Rasmus is wandering about thinking about raping others, while Peter is not. As stipulated before, the cognition is not acted upon by Rasmus, so for an observer, they are identical. Yet, intuitively, Rasmus is a morally worse person than Peter. For Director (2022), the only explanation would be that Rasmus endorses a morally wrong idea (p. 429).

This would presage that the propositional contents of a cognition are a defining factor for moral evaluation. Cognitions, or rather their subject, would then not be inherently considered morally wrongful if thought about. Instead, how the agent interacts with the cognition is of importance. For example, thinking about rape would not per se be wrong. One could be writing a bachelor's thesis about cognitions such as rape fantasies. It would then be permissible since the agent does not necessarily need to endorse the object. Although if the cognition was rather a fantasy where you rape your neighbour, then it intuitively appears to be wrongful. How an agent engages with cognition appears detrimental to its deontic status. Endorsement is the key to establishing how cognitions can be morally relevant.

Still, one might object to the idea of endorsement. One could hold, as Sher does, the view of doxastic involuntarism. If cognitions are not within one's control, then how could one endorse what they are about? Section 2.1 would appear to show that there is a case to be made for cognitions being within one's control. If one were not in control of cognitions, then bad cognitions, such as Islamophobia or rape, would be nullified: one cannot be blamed for having something one cannot help having (Alston, 1988, p. 290). Howbeit, suppose that one was in full control of one's cognitions. If one then chose to have these bad cognitions, then it would

be wrong. Since the question of control appears to nullify responsibility, it implies that there is something *prima facie* wrong with endorsing morally wrongful objects (Director, 2022).

Another aspect that would be useful to elaborate on is the phenomenology of endorsement, since it is the precise transition from passive mental activity to active agency. This can be further clarified by Katharina Anna Sodoma's (2023) definition of endorsement as a positive higher-order mental state aimed at a first-order mental state, such as anger toward a neighbour (p. 173). It is the act of giving one's approval to a cognition, which can lead to cognitive consonance or harmony. There is a reverse of this concept, called *self-distance,* a negative higher-order mental state directed towards a first-order mental state. This would support the notion that it is not always incorrect to engage with any morally laudable content. If Rasmus in ATW instead opted to self-distance from the cognition, then he is not endorsing it. Nevertheless, given the endorsement, he failed to manifest affective critical self-distance from it. Thus, he appears to be morally condemnable.

Another distinction that could be made is between the degree of endorsement. Agents could display *strong* and *moderate* versions of affective endorsement (Jäger & Bänninger-Huber, 2015, pp. 807–808). Agents provide moderate endorsement as long as they do not experience a negative higher-order mental state about their first-order mental state; if agents, however, experience a positive higher-order mental state about this mental state, they are displaying a strong affective endorsement. This would implicate passivity as relevant; even if an agent does not have a positive higher-order mental state, they are still endorsing a moral object. For example, if James hears about a plan to bomb a school, which would kill many children, and does not gain a positive higher-order mental state from it, such as pride, then James would still be endorsing it, although moderately. If James were instead to gain a negative higher-order mental state about it, such as guilt, shame or the feeling that the act is morally laudable, then he would instead self-distance himself and would *not* be endorsing it. This indicates that agents can hold varied amounts of endorsements for an object, with subsequent implications for how they can be judged.

If this were to be true, then any endorsement one makes that is within one's voluntary control could be morally evaluated. Even if many cognitions are not under direct control, they may be under indirect control. To the extent that Rasmus' endorsements are the result of his indirect control, then he can be blamed for their endorsement of those beliefs (Director, 2022, p. 431).

2.4 The Case for Cognitions as Morally Relevant

It appears that one can exert control over some of one's cognitions. Since an agent has the ability to have control of some cognitions, then they and those cognitions are subject to moral evaluation. Even if some cognitions are spontaneous rather than deliberate, one can still have some implicit intentionality over them, showing that it is within one's control.

To follow up on Section 2.2, I hold that the difficulty of the demarcation of cognitions could be sidestepped. By engaging with a cognition, an agent can endorse it or not. If an agent endorses a morally questionable or even wrong object, then it is reasonable to hold that it is subject to moral evaluation. If they do not endorse it, such as when one reads about Nazi atrocities in a history book, and one imagines the scenarios with horror, then there is no case of endorsement, and hence not relevant. The demarcation could still be relevant, but it may not be worthwhile. It is not what has occurred prior to or after the cognition (Basu & Schroeder, 2018). It is the cognition *qua* cognition that is wrong. The same as when one kills one's neighbour; it is the act *qua* act that is wrong.

So far, the intention has been to establish that agents can be held morally responsible for certain cognitions, akin to someone uttering certain phrases, such as racial slurs or misogynistic comments. These cognitions, or their object, can themselves be wrong, but for an agent to be said to be wrong by having them, then they have to endorse them, since just because an agent thinks of these topics does not necessarily follow that their cognitions are morally relevant.[12] If an agent endorses the cognition, it seems morally relevant. Thus, it appears that purely private cognition can be morally relevant, albeit it depends on its propositional content and how an agent interacts with it.

An aspect that was briefly discussed in Section 2 is whether truth or reality factors into whether a cognition is wrong. If someone fantasises about inflicting pain upon a real person, such as your neighbour, then it would be wrongful. But what if it was rather about a fictional character, such as the Phantom? Would that be wrong? One cannot *actually* manifest what one's cognition was about, since the character does not exist, so it would not necessarily follow that it is wrong. To quickly deviate to a similar topic, is the fetishisation of rape or sadomasochism wrongful? The agent does not then base their cognitions on reality, but rather upon a fantasy, or rather a simulation. It is not, per se, a replication of the acts; they have different structures, with the most prominent being that the simulation lacks the wrong-making

[12] This does not mean that cognitions are amoral or that their moral status is agent-dependent. Constructs such as xenophobia, eugenics and misogyny are morally wrongful due to their content.

qualities of the actual event (Hopkins, 1994, p. 124). The activity of fantasising about rape, and not acting it out, is devoid of the qualities of causing harm, limiting freedom, etc. An agent is attending the simulation *qua* simulation, with the act functioning as a substitution for the actual event.

Yet, as John Corvino (2002) argues, this is a wrongful conception of the psychological nature of fantasies. A simulation is not the *object* of desire, but rather the *vehicle* for it. It is not the depiction of the event that is of interest for an agent, but rather what is depicted. A film critic may attend to a film as a simulation (assessing the acting or the production design), but a viewer may be taken with what is depicted rather than the film as a depiction (Corvino, 2002). Agents do not interact with simulation *qua* simulation. By holding *pro-attitudes* toward rape, bigotry and abuse, agents are engaging with and *endorsing* its content. It fails to acknowledge its inherent, wrongful nature. Returning to whether reality factors into cognitions, it seems that it does not. If an agent endorses wrongful objects or activities, then it is rightfully subject to moral judgment and disapproval. The fictional nature of fantasies and their content is not pertinent to whether or not it is right or wrong to have them. Still, intuitively, it feels *prima facie* worse to fantasise about killing your neighbour than the Phantom. Accordingly, if the nature of a cognition is of reality, then it would weigh more than if it were fictional. It may be wrong to think of killing the Phantom, but it is worse to think of killing your neighbour.

So, for a cognition to be morally relevant, it must include a representational state that is directed and involves psychological agency (Basu & Schroeder, 2018).

For ATW to be morally relevant, it needs to be directed, involve the endorsement of a moral object and be subject to the agent's psychological agency over it. The cognition that Rasmus has in ATW is directed (see Footnote 2) and certainly endorses the concept of rape. The agent has psychological agency over cognitions, either directly choosing to have the cognition or by implicitly manifesting intentionality. [13] If one looks back at ATW, the conclusion that ensues is that it is morally relevant and likely condemnable. But if cognitions are said to be wrongful, what does this wronging entail?

3. Cognitions as Morally Wrongful

The idea that one can wrong another by virtue of beliefs is called doxastic wronging (Basu & Schroeder, 2018). There are three important aspects of this construct. Firstly, a doxastic wrong

[13] It has not been discussed whether this is a common occurrence for Rasmus, as the conclusion could be skewed by the habit rather than the cognition *qua* cognition. It would rather be an aretaic issue then. The attempt is to establish that having the cognition, whether it is the first time or the hundredth time, is wrong.

is directed. If you wrong someone, you do wrong to them; you do not merely do wrong. Secondly, they are committed by beliefs. It is not of normative relevance what occurred prior to, or subsequent to, forming the belief; it is the belief itself that is influential. And finally, doxastic wrongs are wrongs in virtue of what is believed. It may be wrong in virtue of its consequences; its effects may have a normative value, but the wrongness lies in the belief and is of greater or all importance. The full extent of how one has wronged cannot be fully understood without addressing a wrong in the belief itself (Basu, 2023, p. 4).

The overlap between doxastic and cognition may be argued to be tenuous, although cognitions can be held to exhibit doxastic wronging. As mentioned, a cognition can be directed; it is not important what occurred prior to or after the cognition; and it is the cognition *qua* cognition that is wrongful, not its consequences. Ergo, it appears that one can consider that cognitions can inflict doxastic wrongs.

The wrong in this case would entail a normative failure through an agent's cognition being prohibited. An act is wrongful when there are reasons that one ought not to perform it (Simester, 2012). A paradigmatic version of this failure is harm. It is wrongful to allow harm to occur. If an act would harm someone (and some alternative act would not), then there is some moral reason against it, writes Johansson and Risberg (2022, p. 509). Agents have a prudential reason to hinder harm, if possible. Harm can function as not respecting a person's moral standing[14], either through physical damage or by not respecting their desires. The latter is the theory of *desire satisfactionism*, which generally posits "that it is bad in itself for people when their desires are frustrated, that is, when what they want to occur fails to occur" (Heathwood, 2026, p. 1). It rejects the *experience requirement* that agents must subjectively experience the harm for it to be considered morally relevant. The only thing that is intrinsically bad for someone is the frustration of their desires. A desire, e.g. wanting respect from one's peers, is frustrated just in case its object does not obtain, or is false. This does not imply that an agent *needs* to experience since it can occur without affecting their experience (p. 9). Someone could be spreading rumours about oneself that make everyone disdain rather than respect them. Though one will never know it, one's desire is not satisfied, but rather frustrated. Agents can then suffer an unfelt desire frustration on this account.

The concept of "harm" is pervasive in moral discussions, though it can be argued that it is insufficiently defined. There are varied conceptions of harm (Bradley, 2012; Hanna, 2016);

[14] The criteria for granting moral standing are disputed, with varied conclusions on the matter (Singer, 1993; Williams, 2006). This paper adopts an existentialist perspective, assuming that entities possessing consciousness have moral standing in virtue of having an irreplaceable personal value to them (Niikawa, 2018).

however, there is a *prima facie* conception of harm that it "somehow involves affecting the subject's well-being adversely" (Johansson & Risberg, 2022, p. 512). From this intuition follows a proposed analysis of harm:

> *Negative Influence on Well-Being Account* (NIWA): What it is for an event *e* to harm an individual *S* is for *e* to adversely affect *S*'s well-being. (p. 510)

This analysis of harm is to be preferred over other forms, due to the others having insufficient explanatory capability for certain events, as well as normative failures.[15] A prime reason is that it is consistent with the importance of harms. Johansson and Risberg (2022) argue that the following principle seems plausible for harm, no matter the exact extent of the harm's normative importance:

> *Normative Reasons* (NR): If *a* and *a** are alternative actions open to *S* in a choice situation, and *a* would harm someone whereas *a** would not, then *S* has some (not necessarily strong or decisive) [normative] reason to perform *a** rather than *a*.[16]

This principle seems reasonable, if not undisputable. It seems evident, Carlson (2020) writes, that one has reason to choose an option that is of benefit to oneself or others, rather than one that is not. Relative to other accounts, NIWA proposes that there exist reasons to prefer more beneficial situations over times harm occurs or when there is no benefit.

From this, it appears that the preferred account of harm is NIWA. Cognitions need to be understood as capable of negatively affecting someone's well-being. Agents would then have NR to avoid normative failures and harmful situations.

3.1 Harming

The discussion of harm is formulated in terms of acts. Hence, it would benefit from being extrapolated to cognitions.

The translation from act to cognition would appear to be rather simple. If an act can be condemned in public form, as in an utterance, then it can be condemned in private form, as a

[15] For a further and more in-depth analysis of this topic, see Johansson and Risberg (2022).

[16] This is adapted from Johansson and Risberg (2022, p. 524) and altered from a prudential to a normative argument. This is to keep the discussion towards morality rather than rationality. The focus should not be on who is harmed, since then it could fall into a self-interest argument (Parfit, 1984), which is not the aim. The rewrite of the argument, replacing "*a* would harm *S*" with "*a* would harm someone", is argued to be justified since the core of the argument is still applicable.

cognition, as well. If an agent were to utter "I want to rape someone", as in ASW, and it is considered condemnable, then the cognition "I want to rape someone", as it was in ATW, would be as well. By condemnation, it means that an agent has performed a normative failure, or rather, that they have wronged.

If wronging is understood as adversely affecting someone by performing a normative failure, then it would be beneficial to establish if and how cognitions can harm. As will be established further down, the harm is unperceived and designated as *unfelt harm*.

Sher (2021) conceptualises harm towards someone as having made "someone less well off in relation to some baseline" (p. 32). Distinct from NIWA, this approach is a version of the *Counterfactual Comparative Account* (CCA) and posits that one is rendered better off if a certain event had not occurred (Hanna, 2016; Klocksiem, 2012). While Sher relies on CCA of harm, this paper will instead utilise NIWA.[17]

Sher's (2021) account of harm could be viewed either through physical well-being or desires. He argues the most direct route is the latter. He continues by writing that "if a person's well-being is reduced whenever his desires are frustrated, then one person can worsen another's simply by thinking of him as he prefers not to be thought of" (p. 32). One is not aware of this, but these cognitions are incongruent with the propositional objects of one's desire. The frustration of desires is an apt account of harm and is congruent with NIWA.

There is a difficulty when it comes to cognitions that are purely private and never publicly manifest, such as ATW, for how could they be argued to adversely affect a person's well-being?

To establish the wrongness of cognitions, it will be argued that they can inflict *unfelt harms*; harming acts that never affect agents' subjective experience (Boonin, 2019; Director, 2022). A prototypical case of it is the following:

> *Henri and Marianne*: Henri is together with their partner, Marianne. They lead normal lives and hold each other in high regard.
>
> *Sean and Joni*: Sean is together with their partner, Joni. They lead a similar life to Henri and Marianne; however, Sean fantasises about killing Joni. Sean will never act upon this, yet still fantasises about it. Joni's subjective experience is never affected: they will never be aware of this.

[17] CCA is one of the other accounts of harms that Johansson and Risberg (2022) dismiss due to technical and normative failures. Why CCA is an inadequate account of harm, see e.g. Folland, 2023; de Villiers-Botha, 2018; Johansson & Risberg, 2022 for further analysis.

Sean and Joni's relationship is outwardly identical to Henri and Marianne's. Yet, it appears *prima facie* wrong to claim that there is no difference between the cases. Intuitively, it feels that things are worse for Joni than for Marianne. Even with the stipulation that it will not affect Joni's experience, it seems *obvious* that Marianne is in a more preferential situation than Joni. This suggests the existence of unfelt harms. If unfelt harms are possible, then Joni is being harmed by their partner's cognitions. If an agent can be harmed without their subjective experience being affected by the disrespect of their moral standing, then unfelt harms could be accepted as genuine harms. And if any kind of harm is *prima facie* wrong, as NIWA indicates, then unfelt harms are as well *prima facie* wrong. And as Sean and Joni's situation would indicate, there appear to exist unfelt harms. As mentioned in Section 3, agents can have propositional desires which are apt to be satisfied or frustrated without any subjective experience of them. Joni may have a propositional desire to be respected and not to be fantasised about killing. Sean's cognition is incongruent with this desire, subsequently constituting an objective worsening of her relationship with the world (Director, 2022). If desires can be frustrated unknowingly, as this would indicate, then this will constitute an unfelt harm towards the subject, and would thus be a genuine harm. Even though Joni's subjective experience will never be affected by Sean, Joni's moral standing is disregarded, and thus, their well-being is adversely affected.

Albeit, there is an objection to this. Chris Heathwood (2022) writes that it can be held that desire frustration can be argued to be a mere *privation*, a lack of good, and not a *robust bad*, a bad in its own right, which is needed to explain how something can be worse than neutral, or rather positively negative, in well-being (p. 3). He holds, however, that this objection does not hold. Consider these two scenarios:

(1) It is bad weather. You lack a desire for it to be good *or* bad.

(2) It is bad weather. You want good weather.[18]

Given that it is bad weather, is it worse to want good weather rather than being indifferent? Intuitively, it appears so. (1) appears to be *well-being neutral*, since it does not involve any desire satisfaction. Meanwhile, (2) seems worse than (1). Thus, it would appear that (2) does contain a robust bad for oneself. For if not, it would be neutral, and only be a mere lack of a

[18] The desire frustration in this scenario may not, in fact, be a true case of a robust bad. Rather, the point of the example is to indicate that desire frustrations are candidates for being robust bads. This would show that it is initially plausible. For further discussion, see Heathwood (2022).

good. The only difference between the scenarios is the presence of an unfulfilled desire. This indicates that they *can* be considered as robust bad.

It appears then that harm does not need to be subjectively experienced for it to wrong someone. From this, it will be argued that the wrong that cognitions are said to be able to produce is an unfelt harm.

3.2 Wrong with Cognitions

The previous section made a case for why cognitions possess the ability to cause harm. It thus seems that cognitions are not different from public acts, such as speech. Cognitions then pose a risk of harm against other persons, similarly to public acts. But then, is it a risk that is permissible? Sher holds that it is, in general, wrong "to risk inflicting a given type of harm only if actually inflicting such a harm would also be wrong" (p. 35). Since risk features normatively in practical reasoning, agents should be concerned with them (Placani, 2017). Yet, for Sher, cognitions possess no risk of harming others. Even inflicting actual frustration of desires would not make an action wrong, and thus, merely risking to frustrate would not follow as well. Risking frustration has no general effect on the moral status of acts, such as cognitions. He writes:

> It is hard to see how we could be, for given both the vast number of people whose desires can be satisfied or frustrated by any person's actions and the variety and heterogeneity of the desires that every person has, even the most ordinary of actions will often frustrate some people's strongly held desires. (Sher, 2021, p. 35)

By this, Sher means that one lacks the responsibility to conform one's cognitions according to others' desires. The author illustrates this through an example where a person buys the last piece of garlic, thereby frustrating the desires of other customers (p. 35). Agents should then not be responsible for risking frustrating others' desires or harming them through cognitions, inasmuch as it cannot even be established for acts. The requirement is not of a moral nature, but rather of intentions (pp. 39–40).

Three points could be raised against this. First, Sher (2021) overlooks the distinction between *incidental* and *directed harm*. This incidental harm is a consequence of one's act and would occur even if an act was performed with good intentions, such as providing for one's starving family. Yet, there is a difference between the incidental harm of an act and the directed wrong of an act, e.g. a cognition. Since cognitions are directed towards a subject, they

necessarily involve them directly. By holding a misogynist belief, one is not *incidentally frustrating* women's propositional desire, but rather *directly harming* them. They are the subject of cognition. It would be as if one bought the last of the garlic to ensure that a specific customer could not buy it. If cognition is not directed, then it cannot be understood to harm others or frustrate their desires. But if it is directed, then it is susceptible to the framework of doxastic wronging.

Second, there are NR against causing harm. An agent may not be able to avoid adversely affecting anyone, but there are still reasons not to inflict harm. If one is in a situation where one has the ability not to cause harm by indulging an alternative action, then there appears to be some NR for performing it.

With the endorsement of a cognition, the agent then opts for a certain action. And by not doing it, they are not adversely affecting someone's well-being. Intuitively, this speaks to there being NR against having certain cognitions, or rather, endorsing certain objects. As persons, we have reasons to respect others as people with a distinct moral standing, and not just hold them as objects (Basu, 2019; Kant, 1785). If an act does not acknowledge this, then there would be a reason against it. Thus, if a cognition, such as the one in ATW, would appear in Rasmus' head, the agent could either opt to endorse it or disregard it, and subsequently either respect or disrespect a person's moral standing. Sher's argument is formulated on doxastic involuntarism and thus does not allow for agents to exert control over their cognitions. But if agents do in fact possess the ability to have volitional control, then it would indicate that they should opt for the alternative action that does not inflict harm. Agents should then have reasons not to harm others.[19]

Regarding the imposition of the risk of harm, one can contend that the intentionality of a cognition is *pro tanto* wrongful, since acting with the intent to cause harm is to act wrongfully, Placani (2017) maintains (p. 83).[20] As argued in Section 2.1, an agent can have implicit intentionality over cognitions. Thus, by engaging and endorsing cognitions, it would then lead to harm by performing a normative failure, such as having a rape fantasy, as in ATW.

[19] Yet, moral reasons against harming may not be so evident. There are cases where the moral reason against impeding a person's well-being is weak at best, with it possible for one to argue a millionaire would not be harmed by being short-changed at the grocery store (Kleinig, 1978). This would suggest that there is not always a *strong* moral reason against causing harm (Gardner, 2021). Still, even though there may not always be strong or even decisive reasons against harming, there will still be some NR against not performing harmful acts.

[20] Placani (2017) does not hold that cognitions alone can harm; it requires a further act for it to be wrong (p. 88). Still, her discussion is apt to be utilised since it adheres to her points around (I) intentionality as the source of moral significance, (II) the requirement of agency, and (III) the culpability of the agent or rather their endorsement of the act. It thus seems applicable.

Attempting to harm, through endorsing a morally wrongful object, is to act wrongfully.[21] This risk of harm is *itself* harmful, argues Placani (2017, p. 86). By disrespecting a person's moral standing, they are worse off than if it had not occurred. It is better that one does not think about killing one's neighbour than if one did, since the neighbour's moral standing has been disrespected. She continues by stating that "being made the target of harm is what is wrong with the *unconsented* intentional imposition of risk of harm onto another person" (p. 87). This imposition constitutes a setback to interests such as having their moral standing respected and not being harmed, and thus, they are wronged. There is an objective worsening of the subject's relationship with the world, which has been altered for the worse by the perpetrator (or thinker). The subject need not be aware of this change; in other words, there is no need for their subjective experience to be affected for the wrong to be relevant.

This would show that there are some NR against acting harmfully, and that the potential risk of inflicting harm is itself harmful.

And finally, by focusing on the difficulty of tracking everyone else's desires, it misses the problem of endorsing morally wrongful objects. It can be an interpersonal issue, but it is also an intrapersonal issue to endorse certain propositions, such as paedophilia or racist beliefs (Corvino, 2002). Agents have some volitional control over their cognitions, so even though it is improbable to satisfy everyone's desire, there is still a duty to attend to oneself (Johansson Werkmäster & Werkmäster, 2024). And still, just because it is difficult not to frustrate anyone's desires, it does not annul responsibility.

To summarise, agents appear to have obligations to respect others' moral standings. They can directly harm others by being the subject of their cognition. The harm that it does, if the cognition is purely private, is not subjectively experienced, but still alters the subject's relation to the world by making them worse off. Yet, the harm is still a genuine harm, which, by harms being *prima facie* wrong, means that it is as well wrong. It is wrong because it has a negative effect on their well-being. Agents are in control of their cognition either through implicit intentionality or active deliberation by endorsing the moral content that the cognition refers to. It appears that cognitions can, in fact, be morally relevant.

[21] The discussion is not concerned with whether an agent has explanatory or justifying reasons for their cognitions. This is why it is only a *pro tanto* wrong.

4. Conclusion

There are a few disparate conceptual threads in this discussion surrounding the moral relevance of cognitions that need to be synthesised.

Cognitions that concern moral objects may not explicitly assert a direct truth-claim. Yet, the propositions' inherent structure expresses truth conditions. The framework of doxastic wronging, proposed by Basu and Schroeder (2018), is also applicable to them.

The mere engagement with these objects may not be wrongful; it is rather *how* it is engaged with that is of importance. Endorsement, as a positive higher-order mental state, of morally wrongful objects, is what is objectionable. The case holds for statements, so why would it not be relevant for private affairs as well? The act of endorsing an object provides intentionality, with cognitions having been established to be apt for control in Section 2.1

This can negatively affect well-being by causing unfelt harms. It objectively worsens a subject's relation to the world and frustrates their desires, as Sean and Joni's situation indicates. Subjects need not be subjectively aware of this alteration, either in relation to the world or the status of their desires, for it to be considered harm.

This can be further understood by returning to ATW. The cognition is wrong since it is (1) directed, (2) the agent exercises agency over the cognition, (3) the agent endorses it, and (4) the endorsement is harmful by disrespecting the moral standing of the subjects, even though they may not subjectively experience this. Rasmus' cognition produces some unfelt harm; in this case, against persons overall. Withal, asserting that this harm is the core of the cognition's wrongfulness would appear tenuous. Instead, it is the endorsement of the moral object of rape that is *prima facie* more objectionable. It is a normative failure by the agent to engage with the object in that manner. One does not merely wrong through word and deed, but also with the intentionality of one's cognitions.

This is not to be understood as advocating for a paternalism against the mental; it is rather to assert that one's private endorsements are acts for which one is morally accountable. The "privacy" of the mind does not grant an exemption; to think is to act, and to act is to be subject to the moral domain.

Th seems thus to be confirmed. By defining the conditions of content, endorsement, and psychological agency, it appears that private mental states are genuine acts subject to normative evaluation. The mind is not, as Sher contends, the wild west.

Bibliography

Alston, W. P. (1988). The Deontological Conception of Epistemic Justification. *Philosophical Perspectives, 2*, pp. 257-299.

Antony, L. (2009). Thinking. In A. Beckermann, B. P. McLaughlin, & S. Walter (Eds.), *The Oxford Handbook of Philosophy of Mind* (pp. 607–630). Oxford: Oxford University Press.

Asay, J. (2021). TRUTH: A Concept Unlike Any Other. *Synthese, 198*(2), pp. 605–630.

Basu, R. (2019). What we epistemically owe to each other. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 176*(4), pp. 915–931.

Basu, R. (2023). The Morality of Belief I: How Beliefs Wrong. *Philosophy Compass, 18*(7), pp. 1-10.

Basu, R., & Schroeder, M. (2018). Doxastic Wronging. In B. Kim, & M. McGrath (Eds.), *Pragmatic Encroachment in Epistemology* (pp. 181-205). New York: Routledge.

Boespflug, M., & Jackson, E. (2024). Doxastic Voluntarism. In E. N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2024 ed.). Stanford: Metaphysics Research Lab, Stanford University.

Boonin, D. (2019). *Dead Wrong: The Ethics of Posthumous Harm.* Oxford: Oxford University Press.

Bradley, B. (2012). Doing Away with Harm. *Philosophy and Phenomenological Research, 85*(2), pp. 390-412.

Carlson, E. (2020). Reply to Klocksiem on the Counterfactual Comparative Account of Harm. *Ethical Theory and Moral Practice, 23*, pp. 407–413.

Corvino, J. (2002). Naughty Fantasies. *Southwest Philosophy Review, 18*(1), pp. 213-220.

de Villiers-Botha, T. (2018). Harm: The counterfactual comparative account, the omission and pre-emption problems, and well-being. *South African Journal of Philosophy, 37*(1), pp. 1-17.

Director, S. (2022, September). The Sheriff in Our Minds: On the Morality of the Mental. *Journal of Ethics and Social Philosophy, 22*(3), pp. 423-437.

Folland, A. (2023). Feit on the normative importance of harm. *Theoria, 89*(2), pp. 176-187.

Gardner, M. (2021). What Is Harming? In J. McMahan, T. Campbell, J. Goodrich, & K. Ramakrishnan, *Principles and Persons: The Legacy of Derek Parfit* (pp. 381 – 395). Oxford: Oxford University Press.

Gendler, T. S. (2011). On the epistemic costs of implicit bias. *Philosophical Studies, 156*(1), pp. 33-63.

Gert, J., & Gert, B. (2025). The Definition of Morality. (E. N. Zalta, & U. Nodelman, Eds.) *The Stanford Encyclopedia of Philosophy*.

Hanna, N. (2016). Harm: Omission, Preemption, Freedom. *Philosophy and Phenomenological Research, 93*(2), pp. 251-273.

Heathwood, C. (2022). Ill-Being for Desire Satisfactionists. *Midwest Studies in Philosophy, 46*, pp. 33-54.

Heathwood, C. (forthcoming 2026). Posthumous Harm. In E. Carlson, J. Johansson, & O. Risberg (Eds.), *The Oxford Handbook of Philosophy of Harm.* Oxford: Oxford University Press.

Heiphetz, L., Landers, C. L., & Van Leeuwen, N. (2021). Does think mean the same thing as believe? Linguistic insights into religious cognition. *Psychology of Religion and Spirituality, 13*(3), pp. 287–297.

Hieronymi, P. (2006). Controlling attitudes. *Pacific Philosophical Quarterly, 87*(1), pp. 45-74.

Hopkins, P. D. (1994). Rethinking Sadomasochism: Feminism, Interpretation, and Simulation. *Hypatia, 9*(1), pp. 116-141.

Jäger, C., & Bänninger-Huber, E. (2015). Looking into meta-emotions. *Synthese, 192*(3), pp. 787-811.
Johansson Werkmäster, M., & Werkmäster, J. (2024). Rejecting norms of standing for private blame. *Ratio, 38*(1), pp. 27–40.
Johansson, J., & Risberg, O. (2022). A Simple Analysis of Harm. *Ergo, 9*, pp. 509-536.
Kant, I. (1785). *Groundwork of the Metaphysics of Morals.* (A. W. Wood, Ed., & A. W. Wood, Trans.) New Haven: Yale University Press.
Kleinig, J. (1978). Crime and the Concept of Harm. *American Philosophical Quarterly, 15*(1), pp. 27-36.
Klocksiem, J. (2012). A Defense of the Counterfactual Comparative Account of Harm. *American Philosophical Quarterly, 49*(4), pp. 285-300.
Kruglanski, A. W. (2004). *The Psychology of Closed Mindedness.* New York: Psychology Press.
Nelson, M. T. (2010). We Have No Positive Epistemic Duties. *Mind, 119*(473), pp. 83–102.
Niikawa, T. (2018). Moral Status and Consciousness. *ANNALS of the University of Bucharest: Philosophy Series, LXVII*(1), pp. 235–257.
Parfit, D. (1984). *Reasons and Persons.* Oxford: Oxford University Press.
Placani, A. (2017). When The Risk of Harm Harms. *Law and Philosophy, 36*(1), pp. 77–100.
Searle, J. R. (1980). The Intentionality of Intention and Action. *Cognitive Science*, pp. 47-70 .
Shah, N. (2002). Clearing Space for Doxastic Voluntarism. *The Monist, 85*(3), pp. 436–445.
Shah, N., & Velleman, J. D. (2005). Doxastic Deliberation. *The Philosophical Review, 114*(4), pp. 497-534.
Sher, G. (2019). A Wild West of the Mind. *Australasian Journal of Philosophy, 97*(3), pp. 483–496.
Sher, G. (2021). *A Wild West of the Mind.* Oxford: Oxford University Press.
Simester, A. P. (2012). Enforcing Morality. In A. Marmor (Ed.), *The Routledge Companion to Philosophy of Law* (pp. 481–494). New York: Routledge.
Singer, P. (1993). *Practical Ethics* (2nd ed.). Cambridge: Cambridge University Press.
Singh, K. (2026). Belief as Commitment to the Truth. In J. Lewis-Jong, & E. Schwitzgebel (Eds.), *The Nature of Belief* (pp. 189–214). Oxford: Oxford University Press.
Sodoma, K. A. (2023). The Significance of Meta-Emotions for Reflecting on Ourselves and Others. *Passion Journal of the European Philosophical Society for the Study of Emotions, 1*(2), pp. 169-184.
Steup, M. (2012). Belief control and intentionality. *Synthese, 188*(2), pp. 145-163.
Willer, D., & Webster, M. (1970). Theoretical Concepts and Observables. *American Sociological Review, 35*(4), pp. 748–757.
Williams, B. (2006). The Human Prejudice. In B. Williams, & A. W. Moore (Ed.), *Philosophy as a Humanistic Discipline* (pp. 135-152). Princeton: Princeton University Press.
Zhou, Y. K. (2022). More on what it means to suffer harm. *Jurisprudence, 13*(4), pp. 516-544.