Stanislav Petrov's Beliefs: A Framework for Doxastic Responsibility

Ryan Park
Stuyvesant High School, New York, USA

Abstract

This essay proposes a framework for when one should be *held* responsible for their beliefs through a case study of Stanislav Petrov, a Soviet officer involved in the 1983 nuclear false alarm incident. I first argue that people are responsible for all their beliefs. I then determine that one believes *responsibly* if and only if the belief is blameless, defending the Blamelessness Thesis over the Praiseworthiness Thesis in the process. This definition of necessary and sufficient conditions for believing responsibly helps elucidate when one should be held responsible for believing irresponsibly. The essay also reconciles doxastic involuntarism, that one does not control what one believes, with doxastic responsibility, and then counters doxastic voluntarism. It concludes by explaining a critical nuance—that people should also be held responsible for lacking certain beliefs.

1. Introduction

It is September 26, 1983, and Soviet Lieutenant Colonel Stanislav Petrov is alerted that a warning system has detected five missiles launched from the United States.[1] Military protocol dictates Petrov must raise the alarm, deferring to superiors who would likely proceed to retaliate. Yet Petrov refuses to believe in the veracity of the reports. Under enormous strain and transgressing protocol, Petrov dismisses the alert as a mistake (as it would ultimately prove).[2] His wise choice prevents probable nuclear war.

For this decision, Petrov is hailed today as "the man who may have saved the world."[3] If so, he clearly believed *responsibly*; though if he had believed and therefore acted otherwise and initiated nuclear war, we would be justified in *holding him responsible* for it.

This essay proposes a framework for when one should be *held* responsible for what one believes. First, I disambiguate the different senses of "responsibility." After doing so, I first argue that people are responsible for all of their beliefs. I find that one should be *held* responsible when believing *irresponsibly*. As a result, I discuss the conditions for believing *responsibly*. I conclude by arguing that people should also be *held* responsible for lacking certain beliefs.

2. Responsible, Responsibly, and Responsibility.

As Rik Peels notes, the word "responsible" has many contextually dependent meanings. In this essay, I will distinguish between (1) believing (*ir)responsibly*, (2) my belief being my *responsibility*, and (3) being *held responsible* for my belief. Our driving question is then to determine the conditions for (3): when we can be held responsible for any beliefs. At a baseline, I take it to be a necessary requirement that the belief one is being *held* responsible for (and, by extension, the holder of the offending belief) is to blame for some damage or harm.[4] But, how and whether to align (1) or (2) with (3) is not immediately obvious. Let us start with what I will call the *Belief-Responsibility Requirement* which aligns (2) and (3): to be *held* responsible for a belief requires that one is *responsible for* said belief. In this light, I ought to be held responsible for my beliefs *if*:

A. My belief (or lack thereof) is the actual or potential cause of harm, and

B. My belief (or lack thereof) is, to some minimally significant extent, my own responsibility.

---

[1] David Hoffman, "I Had a Funny Feeling in My Gut," *Washington Post*, February 10, 1999, https://www.washingtonpost.com/wp-srv/inatl/longterm/coldwar/shatter021099b.htm.

[2] Pavel Aksenov, "Stanislav Petrov: The Man Who May Have Saved the World," *BBC*, September 26, 2013, https://www.bbc.com/news/world-europe-24280831.

[3] Aksenov, "Stanislav Petrov."

[4] Rik Peels, *Responsible Belief: A Theory in Ethics and Epistemology* (Oxford University Press, 2017).

3. Doxastic Involuntarism

Typically, we believe that one can only be responsible for actions within their voluntary control. Doxastic involuntarism, however, is precisely the position that we *cannot* control what we believe. Using William Alston's example, one cannot simply choose to sincerely believe that one plus one is three.[5] Similarly, Pamela Hieronymi argues that it is impossible for one to simply will to believe something even if one believe that it would be good to do so, whereas it is possible for one to physically do something because one believes that it would be good to do so (as with exercise).[6] Note here that even if a hard determinism turns out to be true, it is *still* the case that there seems to be something phenomenologically different between the control we seem to exert over actions and beliefs.

But, if the involuntarist thesis is true, it seems *prima facie* incompatible with *responsibility for* belief, and therefore with being *held* responsible for belief. In other words, in order to secure our responsibility for our beliefs, something cannot wholly be correct in the involuntarist picture.

First, consider the prospects of refuting involuntarism *tout court*. Several substantial arguments have been made in defense of a voluntarist model in which one has some measure of control over one's beliefs. Matthias Steup's "Up-To-Me-Ness" model, one of the most influential of these contributions, presents belief formation as consisting of three moments: (i) an indeterministic state during which one is in the process of weighing at least two possibilities, (ii) the moment one decides which possibility to believe, and (iii) when one starts believing the chosen proposition and cannot voluntarily cease doing so.[7] The voluntarist argument relies on the indeterministic state designated by part (i): at that moment, it is apparently up to the believer to decide between possibilities, which is what s/he does in (ii). This decision implies doxastic control.

But this reliance is not well-founded. The extent to which one actually controls what to believe in (i) is minimal to the point of insignificance because one will automatically begin involuntarily internalizing a judgment regarding it in (ii). Indeed, just because one theoretically has control over what to believe in (i) does not mean in practice that this even counts as control—because no continuation of doxastic control proceeds from it, and even the preceding "free" choice between possibilities will be influenced by external factors and emotional pressures beyond one's voluntary control.

This, however, does not spell the end of doxastic responsibility. It may be possible that even if the involuntarist is correct in that our beliefs are involuntary, we may *still* end up able to

---

[5] William P. Alston, *Epistemic Justification: Essays in the Theory of Knowledge* (Cornell University Press, 1989).

[6] Pamela Hieronymi, "Responsibility for believing," *Synthese* 161, no. 3 (2008): 357–373, https://doi.org/10.11229-006-9089-x.

[7] Matthias Stepup, "Doxastic Voluntarism and Up-To-Me-Ness," *International Journal of Philosophical Studies* 26, no. 4 (2018): 611–18, https://doi.org/10.1080/09672559.2018.1511148.

be held responsible for them. Hieronymi challenges the presumption that voluntary control is a prerequisite for responsibility—instead, she considers doxastic involuntarism evidence that one *should*, in fact, be held responsible for one's beliefs. Because one's beliefs are involuntary, the way one believes provides the truest reflection of one's moral character, which is precisely what we are judging[8].

With this in mind, the *Belief-Responsibility Requirement* cannot do. Rather than aligning *responsibility for* any particular belief as a necessary requirement for being *held* responsible, we instead hold one responsible precisely if they were holding beliefs *irresponsibly*. In other words, aligning (1) with (3). Let us call this the *Responsible-Belief Requirement*: I ought to be held liable for my beliefs *iff*:

A. My belief (or lack thereof) is the actual or potential cause of harm, and

C. My belief (or lack thereof) was held, to some minimum extent, irresponsibly.

Returning to our example of Stanislav Petrov for an illustration, if Petrov had believed that the United States was legitimately attacking and dutifully triggered nuclear war, it would subsequently be reasonable to *hold him responsible* for starting it, even if his judgment of evidence was beyond his control and led him to believe through no voluntary effort of his own. We are still justified in holding Petrov accountable because he had several reasons to dismiss the alert—satellite radio operators had told him they had detected no missiles; he had been trained to expect an all-out attack, not just five missiles; and he had known that the missile detection system was rushed in development.[9,10] Had he raised the alarm, it would have revealed him to be impulsive and uncritical, character traits that are fair to hold one responsible for possessing. Under these conditions, Petrov would have been held responsible for believing that the United States had launched the missiles, as this hypothetical belief would have (A) been the actual or potential cause of harm and (C) been held *irresponsibly*. This argument thus reconciles doxastic involuntarism with doxastic responsibility.

4. Responsible Belief is Blameless Belief

So far, however, it remains unclear what believing responsibly or irresponsibly entails. To fully determine whether one should be held responsible for one's beliefs, it is first necessary to clarify the conditions when one believes responsibly. The most convincing definition for responsible belief is that it is equivalent to blameless belief (the Blamelessness Thesis, or BT), which argues

[8] Pamela Hieronymi, "Responsibility for believing."
[9] Aksenov, "Stanislav Petrov."
[10] David Hoffman, "Funny Feeling in My Gut."

one believes responsibly *iff* one's belief is blameless. Its main opponents contend that responsible belief is praiseworthy belief (or PT), where one only does so *iff* one's belief is praiseworthy. PT is ostensibly supported by Brian Weatherson's example of a cricket captain who is praised for coming up with imaginative field placements.[11] The captain's belief comprises two components: one, that these particular placements are possible; and two, that these particular placements will be successful. This situation appears to support PT because it demonstrates how PT reconciles doxastic responsibility with doxastic involuntarism: the captain seems to deserve more praise for believing and coming up with such placements than another captain who did not, even though the former did not choose to come up with them, (one can never choose to become so inspired). Conversely, it appears that BT fails to capture this nuance; if both captains are blameless (one cannot apparently blame the other captain for failing to come up with equally imaginative placements on the spot), it would follow that both captains believed equally responsibly and therefore deserve equal praise.[12]

However, this application of PT ultimately fails to convince, as the captain does indeed have some measure of responsibility over his otherwise spontaneous beliefs in inspired placements. Though these beliefs are dependent on the contingencies of the game, at all times leading up to it, the captain can choose to spend his time watching as many matches and holding as many practices as possible. This free conduct in one's spare time, committed in these particular ways, would increase the likelihood of imaginative field placement during the game. Therefore, the captain's imaginative placements would reflect, to an extent, his past efforts in training his imagination, which is fair to praise him for.[13] In comparison, it would be fair to blame the other captain for failing to adequately train his imagination for future cricket placements. The result of this argument is that we have good reason to hold the other captain responsible through BT, reconciling BT with doxastic involuntarism.

This situation recalls that used to illustrate Hieronymi's argument, where Petrov should have been held responsible if he had raised the alarm even if he chose to believe, through no voluntary input of his own, that the Soviet Union was under attack. Just as the cricket captain who fails to come up with imaginative placements should be held responsible because he failed to train his imagination, so Petrov should have been held responsible had he believed in the warnings because he would have failed to exercise critical thinking, and his belief would be irresponsibly held, demonstrating an inadequate training for false alarms.

5. Intellectual Obligations

[11] Brian Weatherson, "Deontology and Descartes's Demon," *The Journal of Philosophy* 105, no. 9 (2008): 540–569, https://doi.org/10.5840/jphil2008105932.

[12] Rik Peels, *Responsible Belief*, 46.

[13] Rik Peels, *Responsible Belief*.

BT requires a definition of what it means to believe blamelessly. I take it, following Peels, that one believes blamelessly when one has not violated any intellectual obligations.[14] Put otherwise, responsible belief (or, alternatively, believing responsibly) is equivalent to holding beliefs that are justified with respect to one's knowledge base, while satisfying intellectual obligations.[15]

Perhaps the most influential conception of intellectual obligation is the Jamesian goal of holding as many true beliefs as possible and as few false beliefs as possible. This goal has three variations: "one aims to believe (i) more true propositions and fewer false propositions, or (ii) more true propositions and an equal number of false propositions, or (iii) an equal number of true propositions and fewer false propositions."[16] Each of these three possibilities aligns with the Jamesian goal in some way, with (ii) rejecting the obligation to reduce the number of false beliefs, (iii) rejecting the obligation to increase the number of true beliefs, and (i) requiring action for both true and false beliefs.

Peels holds (i) and (ii) to be unreasonable because any obligation that requires holding more true beliefs would entail a situation in which everyone is constantly violating their obligations, as people would be expected to learn more about everything all the time, which defeats the purpose of specific obligations. If this were the case, Petrov would have violated his obligations to learn more not only regarding matters directly relevant to the nuclear alert situation but also everything else, from chess to ancient Mesoamerican warfare. Instead, Peels argues that a more realistic obligation is (iii). This would entail a moral universe in which no one has any obligation, by virtue of being human, to learn anything about anything. Instead, we would only be responsible for polishing existing beliefs, minimizing any residual errors therein, without any obligation to acquire positively new knowledge and beliefs.

However, it also seems to be the case that one can be held responsible for a lack of certain beliefs. For instance, in democratic societies, it seems reasonable that citizens have *some* obligation to be informed on politics and current events to both sharpen existing positions and acquire new ones in order to better decide on courses of action that have the potential to drastically affect many people's lives, with some actions objectively better than others. For instance, if we wish to avoid the continued devastation of the planet by climate change, citizens have an obligation to learn about it and related matters, formulate and hold beliefs relating to them responsibly (which will then promote responsible action like recycling), and vote for politicians and policies that are likely to reduce its harmful impact. If this obligation did not exist, a future in which billions suffer from the avoidable effects of climate change would be possible without the entire population violating any obligations.

If this is the case, however, Peels total rejection of (i) and (ii) is not wholly correct: we *can*

[14] Rik Peels, *Responsible Belief*.
[15] Rik Peels, *Responsible Belief*.
[16] Rik Peels, *Responsible Belief*, 102.

be obligated to increase the number of true beliefs we hold, in addition to weeding out false/incorrect beliefs. Thus, a revised version of intellectual obligations that escape's Peel's critique is in order. Rather than being obligated to increase true beliefs point-blank, one may have some minimum requirement of true beliefs, or, alternatively, a domain-specific obligation to increasing true beliefs—corresponding to more general deontological duties. Thus, for the citizen, it may be the case that one has a civic duty to stay (somewhat) informed of current events, while for the sports captain, it may be a duty to stay highly informed on sports strategies (by regularly watching film, reading, etc.), as a result of his duty *qua* captain to lead his team. Responsibly believing, then, becomes a matter of satisfying one's existing duties insofar as they contain doxastic elements, and a general duty to remove false beliefs.

6. Conclusion

If Petrov had believed, against his better judgment but with no voluntary input, that the United States had indeed launched nuclear missiles, causing him to raise the alarm and leading to nuclear war, it would be reasonable to hold him responsible for this destruction and holding belief that caused it. Although involuntary, Petrov's hypothetical belief that the United States had launched missiles, would be irresponsible, precisely because holding it demonstrates a violation of his doxastic obligations (per BT) to remain informed about his instruments, and it would have been the cause of immeasurable harm by causing nuclear war. The framework developed throughout this essay provides justification for this conclusion.

References

Aksenov, Pavel. "Stanislav Petrov: The Man Who May Have Saved the World." *BBC*, September 26, 2013. https://www.bbc.com/news/world-europe-24280831.

Alston, William P. *Epistemic Justification: Essays in the Theory of Knowledge*. Cornell University Press, 1989.

Hieronymi, Pamela. "Responsibility for believing." *Synthese* 161, no. 3 (2008): 357–373. https://doi.org/10.1007/s11229-006-9089-x

Hoffman, David. "I Had a Funny Feeling in My Gut." *Washington Post*, February 10, 1999. https://www.washingtonpost.com/wp-srv/inatl/longterm/coldwar/shatter021099b.htm

Peels, Rik. *Responsible Belief: A Theory in Ethics and Epistemology*. Oxford University Press, 2017.

Steup, Matthias. "Doxastic Voluntarism and Up-To-Me-Ness." *International Journal of Philosophical Studies* 26, no. 4 (2018): 611–18. https://doi.org/10.1080/09672559.2018.1511148.

Weatherson, Brian. "Deontology and Descartes's Demon." *The Journal of Philosophy* 105, no. 9 (2008): 540–569. https://doi.org/10.5840/jphil200810