

THOUGHT EXPERIMENTS, ETHICS AND EMOTIONS

EXPERIMENTOS MENTAIS, ÉTICA E EMOÇÕES

Roberto Schmitz Nitsche¹

Abstract: This article aims to analyze the functioning of thought experiments in ethics. We will demonstrate that although many thought experiments in ethics can be reconstructed as arguments, according to the thesis defended by John Norton, its reconstructed version does not have the same epistemic force. We will argue that the reconstructed versions are not capable of bringing the same type of understanding in relation to thought experiments, which have an explanatory advantage that arguments do not have. Therefore, arguments are not a sufficient condition for solving problems within ethics and also within thought experiments in this area. We will argue that emotions can be one of the non-argumentative factors that can produce negative and positive interference on thought experimenters and thought experiment, which cannot be demonstrated accurately in the reconstructed version. In this way, the reconstructed version is not able to convey an adequate understanding of emotions and other influences. Finally, we will elaborate on two thought experiments that demonstrate the need to use emotions in moral actions based on the criticism produced by Flávio Williges against traditional moral theories. These thought experiments can be reconstructed as arguments, but without giving us the same understanding that thought experiments are capable of.

Keywords: Thought Experiments. Epistemology. Ethics. Emotions.

Resumo: Este artigo pretende analisar o funcionamento dos experimentos mentais na ética. Demonstraremos que embora muitos experimentos mentais na ética possam ser reconstruídos como argumentos, de acordo com a tese defendida por John Norton, a sua versão reconstruída não possui a mesma força epistêmica. Argumentaremos que as versões reconstruídas não são capazes de trazer o mesmo tipo de compreensão em relação aos experimentos mentais, que possuem uma vantagem explicativa que argumentos não possuem. Portanto, argumentos não são condição suficiente para a resolução de problemas dentro da ética e igualmente dentro dos experimentos mentais dessa área. Defenderemos que as emoções podem ser um dos fatores não argumentativos que podem produzir interferências negativas e positivas ao experimentador mental e ao experimento mental, que não podem ser demonstradas com exatidão na versão reconstruída. Dessa maneira, a versão reconstruída não é capaz de transmitir uma compreensão adequada de emoções e outras influências. Por fim, elaboraremos dois experimentos mentais que demonstram a necessidade do uso das emoções nas ações morais a partir da crítica produzida por Flávio Williges contra as teorias morais tradicionais. Esses experimentos mentais podem ser reconstruídos como argumentos, mas sem nos trazer a mesma compreensão que os experimentos mentais são capazes.

Palavras-chave: Experimentos Mentais. Epistemologia. Ética. Emoções.

¹ Mestrando em filosofia pela Universidade Federal de Santa Maria, bolsista CAPES.

Introduction

For John Norton, thought experiments are distinguished from other similar things by at least two main characteristics: “[t]hought experiments are arguments which posit hypothetical or counterfactual states of affairs and invoke particulars irrelevant to the generality of the conclusion” (NORTON, 1991, p. 129). It is the use of counterfactuals that gives thought experiments their thought characteristic, without the use of such devices we would only be describing an actual physical experiment or state of affairs. It is the presence of particularities irrelevant to the generalization of the conclusion that gives the thought experiment its characteristic of experiment.

The author argues that all thought experiments can be reconstructed as arguments, and therefore “a good thought experiment is a good argument; a bad thought experiment is a bad argument” (NORTON, 1991, p. 131). By reconstructing the thought experiment in the form of an argument, we can eliminate all particulars irrelevant to the conclusion. This reconstruction, in turn, will not be a thought experiment precisely because it does not have the irrelevant particularities. Norton argues that “[t]hus any conclusion reached by a good thought experiment will also be demonstrable by an argument which does not contain these particulars and therefore is not a thought experiment” (NORTON, 1991, p. 131).

In the first section of this paper, we will analyze whether Norton’s thesis can be applied to thought experiments in ethics; in the second section, we will analyze how the inclinations, for example, racism, culture and emotions of thought experimenters can negatively influence the results achieved by thought experiments; in the third section, we will see how emotions, now seeking a positive meaning, and thought experiments have been used in criticizing traditional moral theories.

1. Thought experiments in ethics cannot be just arguments

Thought experiments in ethics are the ones that probably present the biggest problem to any attempt to produce a universal epistemology² of thought experiments, that is, not only to produce an epistemology that accounts for its functioning in the sciences,

² Such an idea has already been suggested by Cooper: “[A]n account of thought experimentation that can encompass all thought experiments, whether ‘philosophical’ or ‘scientific’, is to be preferred” (COOPER, 2005, p. 229-330).

but of all thought experiments of all areas. When we analyze thought experiments in ethics from the epistemological perspective produced by John Norton, we see that his thesis presents great difficulty in dealing with thought experiments in this area, that is, outside of science. In the same way, we will see that his thesis is not comprehensive enough to explain only in an argumentative way the influence of emotions and feelings in ethical thought experiments. Norton makes it clear that his concern is only related to thought experiments in the sciences, however, we cannot accept such a narrow epistemological explanation³; a good theory must be able to account for the greatest number of cases and, in this sense, Norton's theory will be at a competitive disadvantage compared to other theories that may account for the functioning of thought experiments in several areas⁴. In applying his theory that thought experiments are mere arguments, we face a profound problem: thought experiments in ethics usually end with a question, that is, they ask us what is the right thing to do in a given situation.

Thought experiments that end with a question are at odds with Norton's theory, as there is no result, that is, a chain of arguments that leads us to a conclusion. One point we can discuss to try to save Norton's theory is what a thought experiment is. In both physical and thought experiments, we cannot call them experiments if they are not aiming at a result or have already done so. When thought experiments end with a question, we are not actually dealing with thought experiments, they are proto-thought-experiments, that is, they are merely a description of the experimental scenario. They are not finished; its objective is precisely that someone reads the description and imagines herself in the scenario and only then concludes what is the right thing to do. In short, we can only call a thought experiment something that produced a result, or that at least intends to produce it⁵. The simple descriptions of scenarios, as in the case where they end in questions, aim

³ According to Cooper, for two reasons: “[f]irst, on grounds of simplicity, if it is possible to produce a unified account of thought experimentation, this should be preferred. [...] Second, there are reasons to be skeptical of the idea that science and philosophy are radically distinct enterprises. The work of empirically inclined philosophers of mind and language is often indistinguishable from work in theoretical psychology or linguistics” (COOPER, 2005, p. 229).

⁴ The theory of models of possible worlds proposed by Rachel Cooper is an example.

⁵ According to an anonymous referee, my view of thought experiments necessarily ending with some kind of conclusion is very restrictive: “several authors contradict this view. To name just one of them: according to the taxonomy proposed by Brown in his 1991 book, the type of thought experiment discarded by the author of this article is perfectly included in his classification which, although recent, has now become canonical.” Many things could be said about this problem pointed out by the anonymous referee; however, it must be clear that the purpose of this article is not to define what thought experiments are or to propose an exhaustive classification about them. I should only point out that none of the thought experiments presented in this article lack a conclusion, they all have some kind of result. What I'm doing here is different from what Brown did in his book, I'm not classifying types of thought experiments, I'm just defining what thought experiments cannot be, that is, they cannot be just scenarios. If that were the case, the number of

to reach other people so that they perform the thought experiments from the given scenario, and then, from their worldviews, produce a result. The same idea can be applied for when we only illustrate a thought situation, hypothetical or counterfactual; we do not produce results, we only describe the scenario.

Even with that answer, Norton's epistemology is also unable to offer answers to another serious problem. If we could solve problems in ethics with simple arguments, we would not be in such big trouble⁶ to be discussed by scholars of ethics, such as questions about abortion, killing, eating meat, torturing, etc. There is something else about ethics that cannot be reduced to a simple discussion of arguments; we can come up with a Trolley-type scenario in which one person being killed would ultimately save the lives of hundreds of people. In that case, many people would probably opt for the death of just one. But if the person to be killed is the mother or father of the thought experimenter, we would no longer have the same number of people opting for the death of one to save many.

We follow Brown's criticism: "Norton is right in one sense [...] when he claims that the picturesque details play no role in the argument: [it] is irrelevant to the derivation of the conclusion. But [it] is not irrelevant to the understanding of that conclusion" (BROWN, 1993, p. 276). Brown's criticism is directed at the way we understand the thought experiment without the irrelevant particularities. They have a fundamental role in helping us to understand what goes on in the thought experiment. According to Brown "[t]he point [...] is not to prove the conclusion hitherto unestablished, but instead to provide us with that elusive thing, insight and understanding" (BROWN, 1993, p. 276).

things that would be called thought experiments would be unnecessarily large. We would have great problems in differentiating thought experiments from dreams and daydreams, for example, but this is a very controversial point and would need a lot of future discussions. On the other hand, the classification proposed by Brown, as famous as it may be, has not become a relevant part of any epistemological theory of thought experiments, perhaps not even in the theory developed by Brown himself (as he explains only the functioning of Platonic thought experiments). Norton already argued in 1993 that "I found this whole discussion [about the taxonomy] to be one of the least satisfactory in [Brown's] book. The taxonomy seems quite arbitrary and I could see no reason for selecting just the destructive/direct-constructive thought experiments as Platonic (NORTON, 1993, p. 36). Norton argued that "thought experiments are arguments which: (i) posit hypothetical or counterfactual states of affairs, and (ii) invoke particulars irrelevant to the generality of the *conclusion*" (NORTON, 1991, p. 129, emphasis added). Norton makes it clear that there is some kind of conclusion to the thought experiment, however, Norton does not make it clear whether the conclusion is necessary. Anyway, we can say that there is a possibility that Norton agrees with my restrictive view. Finally, if it is the case that what I call here proto-thought-experiments are actually thought experiments, it will in no way affect the other arguments contained in this article.

⁶ As quoted for example by Eric Luis Uhlmann, David A. Pizarro, David Tannenbaum and Peter H. Ditto in the following passage: "[t]he debate between these two camps [consequentialism and deontology] has generated a number of well-known thought experiments where the two broad principles are pitted against each other in one moral decision (e.g., the 'Trolley' and 'Footbridge' dilemmas)" (UHLMANN, E. L.; PIZARRO, D. A.; TANNENBAUM, D.; DITTO, P. H., 2009, p. 480).

In this sense, thought experiments in ethics have many particularities that for Norton are irrelevant, but which in fact are fundamental to the understanding of the thought experiment. In the third section, we will analyze this point in more detail through an example.

Another point that is defended by Norton concerns the epistemic strength of thought experiments and their respective versions in the form of arguments: “[...] all there is to learn about a thought experiment’s epistemic power can be recovered from considering it as an argument” (Norton 2004b, 55). If that were the case, then why do we use thought experiments instead of arguments in many situations? This point is very clear in the area of ethics, where the strength of theories is often demonstrated through thought experiments, not just arguments.

2. Emotions as obstacles to thought experiments

In the next section, we will look at the positive role that emotions play presented in Flávio Williges’ thesis. In this section, however, we will analyze the negative roles that emotions and other influences bring to thought experiments. Our concern here is not about how emotions work within the thought experiment, in the sense that they are an integral part of the experiment, that is, as a fundamental variable to achieve the result of the thought experiment. Our concern, more precisely, is about how emotions can influence the thought experimenter, that is, our concern here is external to the thought experiment, not internal. We know that some feelings are totally harmful when it comes to our moral judgments: we can treat people in different ways because they are of other nationalities, religions, colors, etc. Especially in the case of thought experiments, we must be systematically concerned with the influences that may negatively affect the thought experimenter.

In a study by Eric Luis Uhlmann, David A. Pizarro, David Tannenbaum and Peter H. Ditto it is shown that racism and nationalism play a very strong role in moral judgments:

[...] college students and community respondents were presented with variations on a traditional moral scenario that asked whether it was permissible to sacrifice one innocent man in order to save a greater number of people. Political liberals, but not relatively more conservative participants, were more likely to endorse consequentialism when the victim had a stereotypically White

American name than when the victim had a stereotypically Black American name. [...] [C]onservatives were more likely to endorse the unintended killing of innocent civilians when Iraqi civilians were killed than when American civilians were killed, while liberals showed no significant effect. [...] [P]articipants primed with patriotism were more likely to endorse consequentialism when Iraqi civilians were killed by American forces than were participants primed with multiculturalism. However, this was not the case when American civilians were killed by Iraqi forces (UHLMANN, E. L.; PIZARRO, D. A.; TANNENBAUM, D.; DITTO, P. H., 2009, p. 479).

Another important study that will help to corroborate our argument was produced by Jonathan M. Weinberg, Shaun Nichols and Stephen Stich. People were introduced to some fictional cases, where they should say whether the person described in the scenario had knowledge or not. This study shows how the cultural issue can also be a determining factor in our judgments. We will bring here, briefly, only one of the studies presented in the article “Normativity and Epistemic Intuitions” produced by them:

One day Charles is suddenly knocked out by a falling rock, and his brain becomes re-wired so that he is always absolutely right whenever he estimates the temperature where he is. Charles is completely unaware that his brain has been altered in this way. A few weeks later, this brain re-wiring leads him to believe that it is 71 degrees in his room. Apart from his estimation, he has no other reasons to think that it is 71 degrees. In fact, it is at that time 71 degrees in his room. Does Charles really know that it was 71 degrees in the room, or does he only believe it? (WEINBERG, J. M.; NICHOLS, S.; STICH, S; 2001, p. 439).

The surprising result was that more than 30% of the experiment participants who were Westerners stated that he really knew what temperature it was, while almost 70% said that he just believed it was such a temperature; among Orientals, just over 10% said that he knew, while slightly less than 90% said that he just believed this to be the temperature.

Problems like those pointed out by the two examples “raises the question of whether such an effect undermines the value of the elicited judgements, or whether we should rather conclude that a trolley scenario which alludes to skin color is badly designed, or that philosophers need to be more careful in running thought experiments” (BRUN, 2018, p. 201).

We believe that these brief examples are sufficient to demonstrate how dependent our judgments on ethical and epistemological issues are on factors such as racism and culture. This is a point that must be taken very carefully into consideration when

producing thought experiments, especially in ethics, which generally ask us what would be the right thing to do in a hypothetical or counterfactual situation. Experiments must be free of inclinations such as racism, the experimenter must be as neutral as possible with regard to her moral judgments, just as physical experiments must have their variables well specified and controlled.

Emotions, in parallel, should receive special care when producing a thought experiment scenario; the experimenter's worldviews based on bias and racism should not influence the outcome of the thought experiment. Racism and nationalism certainly cannot be classified as emotions, however, they are examples of how the thought experimenter's worldview and interpretation can negatively affect the results of a thought experiment, because, "in fact, some types of feelings are harmful to guide choices. Racial hatred, for example, can make it impossible to recognize the qualities of the group affected by the emotion" (WILLIGES, 2016, p. 326).

3. Thought experiments and emotions in criticisms about the traditional moral theories

Traditional moral theories have always tended to place great importance on what are the best reasons for taking a particular action. In this line of thought, emotions do not play a fundamental role in making choices, they are more of a "type of reason for action" (WILLIGES, 2016, p. 323). Therefore, we can treat emotions as an irrelevant feature, in Norton's sense, for traditional moral theories. However, as we will see, they are important for the morality of the action, and, consequently, they are important for the understanding of thought experiments that use emotions to criticize traditional moral theories.

Williges points out that "the right moral decisions require [...] an appropriate emotional configuration" (2016, p. 324). The author does not mean that we do not need reasons, on the contrary, they are essential to guide our actions, however, "emotions can also effectively broaden and deepen our understanding of background elements that inform moral choices" (WILLIGES, 2016, p. 324). Often emotions end up bringing a negative role to our actions, "being misleading, blinding us or making us see something that doesn't exist" (WILLIGES, 2016, p. 324), however, the author argues that they are not always "gross occurrences, devoid of cognitive components, but complex processes involving different psychological states" (WILLIGES, 2016, p. 324). Thus, emotions are

important and necessary, but it would be a mistake to claim that the best actions are the result of their use alone, without the element of reason.

Williges presents an example to defend “that the beneficent action without benevolent motivation is morally deficient” (2016, p. 325). We will try to reorganize the author’s example⁷ in the form of a thought experiment, whose objective is to demonstrate that the rational choice method is not sufficient for us to say that the subject practices a truly moral action, even though he can “guarantee the best choice or decision between possible actions” (WILLIGES, 2016, p. 325). Imagine that a person needs some kind of help, like helping a sick person to feed himself, in which we use love or compassion to take care of that person. Let us imagine yet another person, under the same conditions, but in which we do not use love or compassion to care for that person. In conclusion, the author points out that it is only with the use of love and compassion that “the requirements of morality are fulfilled” (WILLIGES, 2016, p. 325).

Another example brought by the author comes from a famous case of conjoined twins. The babies, known as Mary and Jodie, were conjoined, having only one heart and one pair of lungs. These organs were found in Jodie’s body and were essential to Mary’s life; their separation would result in Mary’s immediate death. On the other hand, if they were not separated, the forecast given by the medical team would be death for both within six months. However, the parents did not authorize the operation because it violated their religious beliefs; they believed that the will of God should be fulfilled. Finally, doctors, through legal proceedings, obtained authorization to carry out the separation even against the parents’ wishes. As predicted, Mary died and Jodie survived.

Imagine two thought experiments related to this case⁸ (again we will transform an example of Williges into a thought experiment format): suppose that the doctor rationally deliberated about the above situation, took into account that the parents’ belief would result in the death of two people and that surgery could at least save one of the babies, and therefore the best thing to do would be to proceed with the operation to separate the twins; imagine the same situation, however, now we will use a doctor who is compassionate with the parents’ situation. This second doctor takes into account all the emotional problems that the situation causes in the parents, he understands the parents’ distress condition and is in solidarity with the situation of loss experienced by the parents; finally, he comes to the same conclusion as the first doctor, even taking an emotional

⁷ Based on passages from page 325 of Williges’ paper.

⁸ Based on passages from page 331 of Williges’ paper.

issue into account. Williges concludes that this “second doctor has a strictly moral sense not covered in the first example” (2016, p. 331). In the first case, without the affective element, the author concludes that the doctor “would not be able to show a fundamental moral aspect of moral beneficent actions: that the pain and humanity of others, as well as their needs, are important to us” (WILLIGES, 2016, p. 331-332). These two thought experiments come to the same conclusion: save at least one of the babies. They could be clearly reconstructed as arguments, following Norton’s thesis, but the arguments would not be able to explain why only in the second case is there a properly moral sense.

In this second case, we have an irrelevant feature that, according to Norton’s thesis, could be eliminated in the reconstruction, since the two thought experiments produce the same result. However, it is only with the help of this irrelevant feature, that is, the doctor’s attention to the family, which we understand that only in the second thought experiment we have “a properly moral sense not covered in the first example” (WILLIGES, 2016, p. 331).

On the other hand, Rachels argues that precisely the presence of feelings, as in the case above, prevents us from reaching the right decision on what to do:

Such feelings are often a sign of moral seriousness and so may be admired. But they can also be an impediment to discovering the truth: when we feel strongly about an issue, it is tempting to assume that we just *know* what the truth must be without even having to consider the arguments on the other side. Unfortunately, however, we cannot rely on our feelings, no matter how powerful they may be. Our feelings may be irrational: they may be nothing but the products of prejudice, selfishness, or cultural conditioning (RACHELS, 2003, p. 11, emphases on the original).

As we have already seen, Williges agrees that emotions often hinder moral decisions, such as racism that may not allow us to see the qualities of a person, as well as the love that can lead us to do irrational things, we can also quote “anger, jealousy, emotional states that involve dispersion or lack of focus, and certain feelings of self-indulgence” (WILLIGES, 2016, p. 326). However, that is not all that can be said about emotions. There are emotions that can help us to broaden and deepen moral deliberations such as “interest and admiration (wonder)” (WILLIGES, 2016, p. 327). In the case of the medical team, only if they are “equipped with the appropriate affective sensitivity, [...] compassionate or endowed with compassion, [do they] understand the sense of loss and the family’s despair for disrespecting their religious creed” (WILLIGES, 2016, p. 330).

Thus, emotions do not always need to be irrational or impulsive, “they can be controlled and voluntary, highlighting relevant moral outlines in the context that underlies the decision” (WILLIGES, 2016, p. 330).

Another thesis defended by Williges is that emotions, in the case of charitable actions, have a moral or normative role. We can organize an experimental scenario to understand the author’s point. Let us imagine two mothers who take care of their children, the first does everything that we expect a mother to do for her son, take care of her and also love her; the second mother does exactly the same as the first, but she does it out of duty, without a feeling of love involved, that is, it is just a technical action. The author concludes that “charity action without emotion excludes humanity from the act of caring” (WILLIGES, 2016, p. 332). Thus, “it does not seem reasonable to believe that we can ‘take care’, ‘help’, ‘attend’, ‘protect’ and carry out other charitable actions in a morally full sense if all these actions are carried out mechanically and at a distance” (WILLIGES, 2016 p. 334).

Conclusion

Although many thought experiments can be reconstructed in the form of arguments, as advocated by John Norton, especially those referring to physics and science in general, they do not have the same epistemic strength; thought experiments bring us a much greater understanding of problems than with the simple use of arguments. Thought experiments in ethics need other elements that cannot simply be reduced to argumentative issues. Ethical problems require more than arguments to reach solutions.

Emotions can play a negative role in thought experiments, such as when the thought experimenter is not in a neutral position (in an ethical sense) in relation to the variables tested in the experiment. We showed studies that demonstrated that our judgments change according to our nationality, whether we are racist or if we belong to a particular culture. However, emotions can also play an important role in creating the scenarios, where feelings of interest and admiration can help to make the thought experiment more reliable. Finally, we demonstrated the possibility of creating thought experiments that include the emotional factor as a criticism of traditional moral theories that do not take emotions to such a high level in the factors of moral actions and judgments.

References

- BROWN, J. R. Why Empiricism Won't Work. In: *Biennial Meeting of The Philosophy of Science Association*, Volume Two: Symposia and Invited Papers. p. 271-279, 1993.
- BRUN, G. Thought Experiments in Ethics. In: *The Routledge Companion to Thought Experiments*. STUART, FEHIGE, BROWN, London and New York: Routledge, 2018.
- COOPER, R. Thought Experiments. In: *Metaphilosophy*, v. 36, n. 3, pp. 328-347, 2005.
- NORTON, J. D. Thought Experiments in Einstein's Work. In: HOROWITZ, T.; MASSEY, G. J. (ed.), *Thought Experiments in Science and Philosophy*. Mariland: Rowman and Littlefield, p. 129-148, 1991.
- _____. Seeing the Laws of Nature. In: *Metascience*, v. 3, p. 33-38, 1993.
- _____. Are Thought Experiments Just What You Thought? In: *Canadian Journal of Philosophy*, v. 26, n. 3, p. 333-366, 1996.
- _____. On Thought Experiments: Is There More to the Argument? In: *Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association, Philosophy of Science*. p. 1139-1151, 2004a.
- _____. Why Thought Experiments Do Not Transcend Empiricism. In: HITCHCOCK, C. (ed.), *Contemporary Debates in the Philosophy of Science*. Blackwell, 2004b, p. 44-66.
- RACHELS, J. *The Elements of Moral Philosophy*. New York: McGrall-Hill Companies, 2003.
- UHLMANN, E. L.; PIZARRO, D. A.; TANNENBAUM, D.; DITTO, P. H. The Motivated Use of Moral Principles. In: *Judgment and Decision Making*. v. 4, n. 6, p. 476-491, 2009.
- WILLIGES, F. O Papel das Emoções Altruístas na Escolha e na Moralidade das Ações Benéficas. In: *Filosofia. Revista da Faculdade de Letras da Universidade do Porto*, Portugal, v. 33, p. 321-336, 2016.
- WEINBERG, J. M.; NICHOLS, S.; STICH, S. Normativity and Epistemic Intuitions. In: *Philosophical Topics*, v. 29, n. 1 e 2, p. 429-460, 2001.

Recebido em: 08/05/2020
Aprovado em: 29/05/2020