Open Access

OBM Neurobiology

Communication

A Neurotransmitter Approach to the Trolley Problem

Daniel Z. Lieberman*, Sara Teichholtz, Brenna R. Emery

George Washington University, School of Medicine and Health Sciences, Department of Psychiatry and Behavioral Sciences, Washington DC, USA; E-Mails: dlieberman@mfa.gwu.edu; steichholtz@gwu.edu; brennarosenberg@email.gwu.edu

* **Correspondence:** Daniel Z. Lieberman; E-Mail: dlieberman@mfa.gwu.edu

Academic Editor: Bart Ellenbroek

OBM Neurobiology
2019, volume 3, issue 2
doi:10.21926/obm.neurobiol.1902030

Received: Feburary 25, 2019
Accepted: June 26, 2019
Published: June 28, 2019

Abstract

Is it ethically permissible to sacrifice the life of one human being in order to save the lives of five others? This question forms the basis of the famous thought experiment called "the trolley problem." Surveys have found that 90 percent of people agree it is permissible when the victim to be sacrificed is described as off in the distance, whereas the opposite result occurs if the victim is described as up close. No consistent ethical principle has been identified that account for these results. In this paper we propose a solution based on the neurobiology of the human brain with regard to processing objects in three-dimensional space. We argue that the solution to this problem is that different neurotransmitters are activated when managing objects that are far away (dopamine) versus up close (e.g., serotonin and oxytocin). These different neurotransmitters tend to make people apply different ethical philosophies: utilitarianism and harm aversion, respectively.

Keywords

Trolley problem; ethical issues; moral reasoning; experimental philosophy; dopamine; extrapersonal space; peripersonal space; utilitarianism; deontology; harm aversion

1. Introduction

Since it was first introduced in 1967 by Philippa Foot [1], the trolley problem has become a well-known thought experiment that has captured the attention of ethicists, psychologists, and cognitive neuroscientists. It appears to elicit a contradiction in ethical reasoning that has not been adequately explained despite decades of research and speculative reasoning [2]. In this paper, we introduce a new approach: a neurobiological solution based on neurotransmitter effects in order to account for the apparent contradiction.

The essence of the problem is the question of whether it is ethically permissible to take the life of one person in order to save five. The first part of the thought experiment is as follows.

> *Imagine you are walking a safe distance from a set of train tracks when you hear the sound of a distant train approaching behind you. The driver has fallen asleep and the train is out of control. Up ahead, you see five men working on the tracks. The conditions of their work do not allow them to escape from the tracks. However, you become aware of a switch next to you. You know that if you pull this switch, the train will be diverted from its current course onto a side track. On the side track there is one lone worker who similarly would not be able to escape the train's path. It is not possible to warn either set of workers in time for them to escape. Is it ethically permissible to pull the switch to save five people by killing one?*

Faced with this question, 90 percent of people surveyed agreed that yes, it was ethically permissible to pull the switch to change the route of the train to cause only one death instead of five [3]. In fact, many people went beyond saying that it was permissible and stated that it was ethically imperative to pull the switch.

The second part of the thought experiment maintains the core decision of whether to trade one life for five, but a change is made in how the trade is brought about.

> *You are standing on a bridge overlooking the same set of train tracks. The same five workers are up ahead when you hear the train approaching from behind you. However, this time there is no switch and no side track. Instead, there is a man next to you on the bridge, and you know that if you push him off, his mass will be enough to slow the train and save the lives of the five workers. Is it ethically permissible to push the man onto the tracks?*

Ninety percent of the respondents surveyed believed it was ethically impermissible to push the man off the bridge. Exactly the opposite response compared to the first scenario, despite the fact that the ethical problem is the same (bringing about the death of one person to save the lives of five).

2. Ethical Principles

How do we account for this difference? We can begin by inferring the ethical principles that are being implicitly followed by the majority of survey respondents in the two scenarios. In the first scenario, in which 90 percent of respondents believed it was appropriate to pull the switch to divert the train, the ethical principle being followed is utilitarianism [4]. This theory states that the best action is the one that maximizes benefits. It is clear that there is greater utility in saving the lives of five people versus sparing the life of one. In this situation the application of utilitarianism might be phrased as, "We must act to minimize loss of life." Utilitarianism is a common approach to problems. Sacrificing the safety of the few for the benefit of the many is seen when firefighters

rush into a burning building, when soldiers are sent to war, and when participants are enrolled in a high-risk clinical trial of an experimental treatment, to name a few examples.

In the second scenario, the push condition, the respondents were implicitly following the ethical principle of deontology, which states that an action must be judged right or wrong without regard to the consequences [5]. Another way of saying this is that the ends do *not* justify the means. An unjust act is not redeemed by a beneficial outcome. Using this approach, the question becomes, is it permissible to push someone off a bridge into the path of an oncoming train. The consequences (saving five lives) are not considered within this ethical framework, and so the answer becomes clear. It is wrong to kill someone by pushing them off a bridge. This ethical philosophy is also called "harm aversion" because the goal is to avoid harming people [6]. This approach might be phrased as, "We cannot intentionally harm people even if others will benefit from their suffering."

A real world example of harm aversion is the informed consent process. Tricking people into enrolling in a high-risk study of a potentially life-saving drug would be appropriate within a utilitarian framework. Any harm experienced by the participants would be outweighed by the potential benefit to thousands of people. However, a harm aversion approach tells us that it is unethical to deceive people about research risks regardless of the benefit to others.

There is merit to both of these ethical philosophies, and there can be disagreement over which is better. But if an individual believes that utilitarianism is better, one would expect a rational person would apply it consistently, and the same is true of harm aversion. Either it is ethically permissible to trade one life for five, or it is not. It is not clear why people apply utilitarianism to the switch scenario and harm aversion to the push.

One philosophical solution that has been proposed is that in the push scenario the intention is to kill the man on the bridge [7]. His death is necessary to save the lives of the five workers. In the switch scenario, on the other hand, the death of the one worker is incidental. He is simply in the wrong place at the wrong time. There is no intention to kill him; we just need to get the train off the main track. This solution has been summarized as *kill versus let die*.

It posits that there is a difference between foreseeing the death of an individual and making his death the goal of your actions. Foot argues, "It is one thing to steer towards someone foreseeing that you will kill him and another to aim at his death as part of your plan." [8] This solution was empirically tested in the following way: The switch scenario was modified to make the side track a loop. If the bystander pulls the switch, the train will be diverted to a side track that loops around and then returns to the main track so the death of the five workers isn't prevented, it's only delayed. However, the presence of a lone worker on the looping track will stop the train just like the man on the bridge who is pushed onto the tracks.

In this modified switch scenario, the death of the lone worker is intended as part of the plan just like in the push scenario. The bystander at the switch intends to kill the man on the looping track in order to save the five on the main track. There is no longer a scenario in which the death of an individual is foreseen but not intended. If the *kill versus let die* solution were correct, this modification should make people less willing to throw the switch. However, when the modified problem was tested, the addition of the loop did not change people's reaction to the problem [9]. Most still said that the switch was ethically permissible, but the push was not. In this form of the thought experiment the *kill versus let die* solution failed.

Another thought experiment used to test *kill versus let die* involved a question about organ transplantation [2]. In the *kill* scenario people were asked if it was ethically permissible to take the life of a person so his organs could be harvested to save the lives of five people in need of transplantation. The *let die* scenario was similar, but rather than killing the prospective organ donor, he is said to be dying of natural causes and life-saving treatment is withheld. A large majority of people surveyed responded that both scenarios were ethically impermissible. Once again, *kill versus let die* failed to explain people's reactions to the trolley problem.

Other philosophical approaches have been offered, but none of them have provided a satisfactory solution. Judith Thomson, who has studied this problem in depth, notes that despite all the work that has been done, "the trolley problem… remains with us." [2] No set of consistent ethical principles has been identified that correctly accounts for people's reactions.

3. The Role of Emotion

More recently, a neurobiological approach to the problem has generated interest. A number of studies have identified anatomical regions in the brain that are involved in moral cognition. They include the prefrontal cortex, anterior cingulate cortex, anterior temporal lobes, superior temporal sulcus, insula, precuneus [10], as well as the amygdala, temporoparietal junction, and posterior cingulate [11].

The philosophical approaches discussed above focus on the role of reason in ethical judgements. An fMRI study took a complementary approach by placing the emphasis on emotion [12]. The investigators hypothesized that the push condition would engage people's emotions in a way the switch condition would not, and that would account for the differences in respondents' willingness to sacrifice one life to save five.

Participants were presented with 60 practical dilemmas; some were more emotional, described as "moral-personal," and some were less emotional, described as "moral-impersonal." The results showed that portions of the medial frontal gyrus, the posterior cingulate gyrus, and the angular gyrus, all of which have been associated with emotional processing, were significantly more active in the moral-personal condition. They also found that areas involved in working memory were less active in the moral-personal condition suggesting that less cognitive processing was occurring.

The authors note that the results of this study are a first step toward approaching a solution to the problem rather than a definitive answer. The experiment confirms one's intuition that pushing someone to their death would cause greater emotional activation than killing someone by pulling a switch, but it doesn't explain why the brain functions this way. Why does distance (close versus far) moderate the emotional response, even when the outcomes of an ethical decision are identical?

4. A Solution Based on the Neurobiology of Three-Dimensional Space

We propose that the answer to this question hinges on the neurotransmitters the brain uses to process objects in three-dimensional space (close versus far) and the consequences of differential activation of those neurotransmitter circuits.

The brain uses different circuits to process information about the environment depending on where an object of attention is located with respect to the body [13]. Space can be divided into

two concentric spheres with the body at the centre. The inner sphere is the peripersonal space. It is the space around a person that is within arm's reach. Beyond that is the extrapersonal space.

Things that are within the peripersonal space are within an individual's control. These are typically things that are possessed by the individual. Interaction with things in the peripersonal space involves experiencing them with the senses, enjoying them, and consuming them. Things that are within the extrapersonal space, by contrast, are things that we do not possess. They are outside of our control. Our interactions with them are more limited. We can desire them, imagine them, and make plans to get them.

It makes intuitive sense that evolution would lead to distinct pathways that process information about the peripersonal and the extrapersonal spaces. With regard to resources necessary for survival and reproduction, there is a fundamental difference between resources that are possessed and those that are needed but not immediately available. The way we interact with something we have is different from how we interact with something we need but do not have. They require different approaches and consequently use different circuits and neurotransmitters.

In addition to a spatial divide, there is also a temporal divide. Things in the peripersonal space are experienced in the present. Things in the extrapersonal space are confined to the future. Because they are out of arm's reach, any actual, physical interaction must be delayed. It may be only the few seconds needed to get a book off a shelf, or it could be the years of effort necessary to obtain results from a clinical trial.

There's one more distinction that is useful to highlight when contrasting how the brain manages information in the different realms of three-dimensional space. Things that are in the peripersonal space are experienced with the senses. We can taste, touch, and smell them. Things in the extrapersonal space, if they are far enough away, are experienced only with the imagination. Food which is desired but not available is an abstract concept as opposed to a real thing. The circuits that are able to think about a meal that doesn't exist are also able to think about other abstract concepts such as math, language, and ideas, such as justice.

The primary neurotransmitter that regulates processing of objects in the extrapersonal space is dopamine [13-16]. Dopamine is most widely known for its role in goal directed activity. It is responsible for triggering desire, anticipation, motivation, and when a goal is imminent, reward [17]. These behaviours are associated with activity in the dopaminergic mesolimbic pathway [18] (Figure 1). This pathway provides the initiative required to pursue resources. Pathology within this circuit can lead to the compulsive drug seeking of an addict [19] or the impulsive pleasure seeking of a bipolar patient experiencing a manic episode [20].

Another dopamine pathway, the mesocortical tract, affects the behaviour of the prefrontal cortex. This circuit allows us to go beyond the immediate desire of the mesolimbic pathway and make longer term plans to maximize resources that require multistep planning and executive functioning [21]. Prefrontal dopamine also plays a role in abstract thinking and complex problem solving [22]. Pathology within the mesocortical circuit can lead to the impulsivity observed in people with attention deficit hyperactivity disorder [23].

There are a number of neurotransmitters that are involved in processing information in the peripersonal space. They include serotonin, norepinephrine, as well as oxytocin, because affiliative relationships are treated as being within the peripersonal space [13]. Affiliative social interactions take place for the purpose of enjoyment and social cohesion, for example relaxing with friends or

family [24]. In addition to providing enjoyment and satisfaction, these circuits are involved in empathy and feelings of emotional connection.

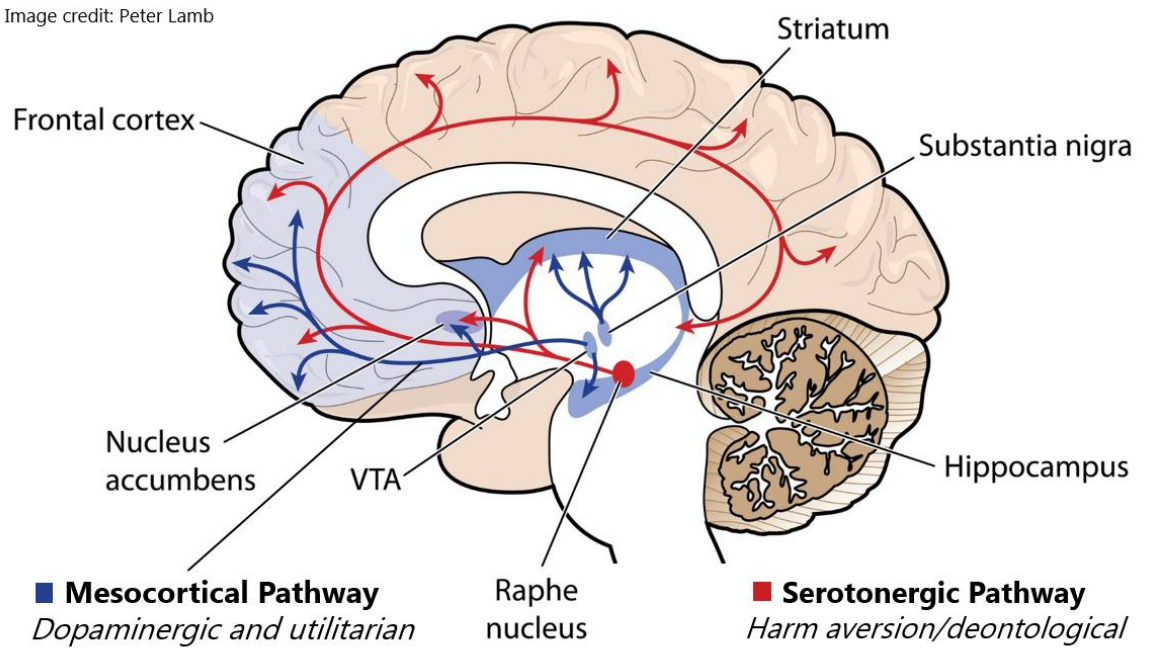

Figure 1 Schematic diagram of the neural pathways of dopamine and serotonin. The dopaminergic mesocortical pathway, which extends from the ventral tegmental area (VTA) to the frontal cortex, is involved in long-term planning, aimed at maximizing future resources. The serotonin pathway has been found to promote here-and-now aversion to inflicting harm on another person.

Alternatively, agentic social relationships are processed by dopaminergic extrapersonal circuits [25]. In contrast to affiliative relationships, which involve immediate enjoyment in the present moment, agentic relationships are for the purpose of accomplishing a future goal, such as cooperating on a project with co-workers, connecting with people at a networking event, or deciding what to do about dinner. Most real-world relationships have both affiliative and agentic components.

We now have the tools to provide a neurotransmitter-based solution to the trolley problem. The switch condition involves the neurotransmitter that processes interactions within the extrapersonal space: dopamine. This neurotransmitter orients behaviour toward an emphasis on maximizing future resources, which is utilitarianism. The push condition takes place within the peripersonal space. It activates neurotransmitters, such as oxytocin, that orient individuals to social connections and stimulate empathy. In a peripersonal situation the attention is focused on the present, and future consequences are given less priority. This leads to a deontological or harm aversion approach.

The connection between the ethical issues of the trolley problem and the way the brain processes three-dimensional space can be made even more clear if we add further scenarios to the problem. In the switch condition, we're still close enough to the victim to perceive him with at least one of our senses, that is, vision. Sensory stimulation is peripersonal in nature, and consequently causes emotional resistance to pulling the switch. If we add additional distance, we further shift towards dopamine and the extrapersonal.

Imagine you are seated in an office on the other side of the country. The phone rings and an agitated person describes the out of control train bearing down on the five workers and the availability of a side track with a single worker. On your desk is the switch to divert the train.

The addition of the extra distance and the elimination of visual stimuli would be expected to make the decision easier. If we add one more component, the component of time, we will squeeze out all of the peripersonal and make it completely extrapersonal/dopaminergic. At that point, the thought experiment no longer presents a dilemma. It becomes easy. Here is how we can take the scenario out of the peripersonal present and shift it to the extrapersonal future.

You are a train safety engineer writing software to control the settings of track switches under emergency conditions. There are cameras on the side of the tracks, and you can write software that will evaluate the potential loss of life in any given situation and automatically program the switches to save the most lives. Is it ethically permissible to write a software program of this nature that might sacrifice a single life to save five?

By replacing a present event with a future event, we are able to maximally suppress the peripersonal circuits that are responsible for harm aversion and rely exclusively on the utilitarian dopamine circuits.

In addition to shifting moral judgement by varying distance and time, moral judgement can also be manipulated pharmacologically. When volunteers were given the selective serotonin reuptake inhibitor citalopram, they became more likely to use a harm aversion approach to moral reasoning. They became less willing to sacrifice an innocent life to save others, and less willing to inflict harm on people as a punishment for unfair behaviour [26]. Overall, the serotonin enhancing drug made volunteers more likely to view harming people as forbidden.

5. Conclusions

The neurotransmitter solution to the trolley problem avoids the assumption that people make decisions based on a consistent set of ethical principles. It replaces a top-down philosophical approach with a bottom-up biological approach. People often make ethical decisions based on an instinctual sense of what feels right rather than using a rational process that relies on the application of consistent rules. Therefore the solution to the problem involves realizing that the expectation of consistent ethical principles by humans is not consistent with the way the human brain functions.

As a result of the circuits the human brain uses to process three-dimensional space, we are more likely to apply harm aversion principles in situations in which we are up close and personal. We are willing to prioritize the well-being of the individual who is close to us over the abstract common good. This helps explain why people who make personal appeals for special treatment, for example to receive an organ transplant, are sometimes bumped to the front of the line, despite the fact that it's contrary to the ethical principle of fairness [27].

Human behaviour is complex and multidetermined. There are other factors that influence moral decision making besides differential activation of neurotransmitter circuits. However, a neurobiological understanding points us toward the conclusion that when we observe from a distance, we are more likely to use a utilitarian approach. This enables leaders, such as politicians, to make difficult decisions that harm people, but are necessary for the greater good, such as taking resources from one group to support another. Activists who oppose these decisions often

try to bring a victim of the decision within the politician's peripersonal space. These activists intuitively recognize that location in three-dimensional space influences moral judgement.

Rules based on logic are not the only rules that humans follow. Human nature is more complex than rational thinking would lead one to believe. Thus, the trolley problem can be seen not only as a thorny problem of ethical contradictions, but also as a key to understanding an important aspect of human nature that follows non-rational rules based on the evolutionary consequences of brain circuit development.

Author Contributions

The three authors all participated in the literature search that provided the background for the trolley problem and in the drafting of this manuscript.

Competing Interests

The authors have declared that no competing interests exist.

References

1. Foot P. The problem of abortion and the doctrine of double effect. Oxford Review. 1967; 5: 5-15.
2. Thomson JJ. Turning the trolley. Philos. Public Aff. 2008; 36: 359-374.
3. Hauser M. Moral minds: How nature designed our universal sense of right and wrong. New York: Ecco/HarperCollins Publishers. 2006.
4. Mill JS. Utilitarianism. Harlow: Longmans, Green and Company; 1985.
5. McNaughton D, Rawling P. Deontology. In: Copp D, editor. The Oxford handbook of ethical theory. Oxford: Oxford University Press. 2006: 459-479.
6. Reynolds CJ, Conway P. Not just bad actions: Affective concern for bad outcomes contributes to moral condemnation of harm in moral dilemmas. Emotion. 2018; 18: 1009-1023.
7. Thomson JJ. Killing, letting die, and the trolley problem. The Monist. 1976; 59: 204-217.
8. Foot P. Virtues and vices and other essays in moral philosophy. Oxford: Oxford University Press on Demand. 2002.
9. Greene J. Solving the trolley problem. In: Sytsma J, Buckwalter W, editors. A companion to experimental philosophy. New York: Wiley Blackwell. 2016; 61:175-178.
10. Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J. The neural basis of human moral cognition. Nat Rev Neurosci. 2005; 6: 799-809.
11. Fede SJ, Kiehl KA. Meta-analysis of the moral brain: patterns of neural engagement assessed using multilevel kernel density analysis. Brain Imaging Behav. 2019; 1-14.
12. Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. Science. 2001; 293: 2105-2108.
13. Previc FH. The neuropsychology of 3-D space. Psychol Bull. 1998; 124: 123-164.
14. Previc FH. The dopaminergic mind in human evolution and history. Cambridge: Cambridge University Press. 2009.
15. Szechtman H, Talangbayan H, Eilam D. Environmental and behavioral components of sensitization induced by the dopamine agonist quinpirole. Behav Pharmacol. 1993; 4: 405-410.

16. Hills TT. Animal foraging and the evolution of goal‐directed cognition. Cogn Sci. 2006; 30: 3-41.
17. Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol. 1998; 80: 1-27.
18. Koob GF. Hedonic valence, dopamine and motivation. Mol Psychiatry. 1996; 1: 186-189.
19. Robinson TE, Berridge KC. The neural basis of drug craving: an incentive-sensitization theory of addiction. Brain Res. 1993; 18: 247-291.
20. Berk M, Dodd S, Kauer‐Sant' anna M, Malhi GS, Bourin M, Kapczinski F, et al. Dopamine dysregulation syndrome: implications for a dopamine hypothesis of bipolar disorder. Acta Psychiatr Scand. 2007; 116: 41-49.
21. Weinberger DR, Berman KF, Chase TN. Mesocortical dopaminergic function and human cognition. Ann N Y Acad Sci. 1988; 537: 330-338.
22. Previc FH. Dopamine and the origins of human intelligence. Brain Cogn. 1999; 41: 299-350.
23. Winstanley CA, Eagle DM, Robbins TW. Behavioral models of impulsivity in relation to ADHD: translation between clinical and preclinical studies. Clin Psychol Rev. 2006; 26: 379-395.
24. Depue RA, Morrone-Strupinsky JV. A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. Behav Brain Sci. 2005; 28: 313-349.
25. Depue RA, Collins PF. Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. Behav Brain Sci. 1999; 22: 491-517.
26. Crockett MJ, Clark L, Hauser MD, Robbins TW. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. P Natl Acad Sci. 2010; 107: 17433-17438.
27. Chambers M. Tough questions about transplants raised by new heart for 'Baby Jesse.' [Internet]. New York: New York Times; 1986 [Accessed 19 Mar. 2019]. Available from: https://www.nytimes.com/1986/06/15/us/tough-questions-about-transplants-raised-by-new-heart-for-baby-jesse.html.