UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Consent and the Doctrine of Double Effect

Permalink

https://escholarship.org/uc/item/3q94615i

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors

Ocampo, Joshua Phillips, Jonathan

Publication Date

2022

Peer reviewed

Consent and the Doctrine of Double Effect

Joshua Ocampo (joshua.c.ocampo.22@dartmouth.edu)

Jonathan Phillips (jonathan.s.phillips@dartmouth.edu)

Program in Cognitive Science, Dartmouth College 23 N Main St. Hanover, NH 03755 USA

Abstract

The doctrine of double effect (DDE) explains that it may be permissible to cause harm as a foreseen side-effect of an action that brings about a good result but impermissible to cause harm as a means of bringing about the same good result. The DDE is commonly illustrated with the Trolley Problem, which along with similarly structured examples, have become widely popular as a tool for studying moral psychology and have been taken to demonstrate a universal feature of moral judgment. Across two studies, we investigate how consenting to being harmed interacts with the Doctrine of Double Effect. Specifically, we ask whether (1) harming someone as a means becomes morally acceptable when that person consents to being used as a means, and (2) whether the distinction between harming as a means vs. side-effect persists even when the person being harmed consents. We find that consent significantly interacts with the DDE.

Keywords: consent; doctrine of double effect; moral psychology

Introduction

On one formulation, the doctrine of double effect (DDE) holds that it may be permissible to cause a harm as a foreseen side-effect of an action that brings about a good result, while it would be impermissible the cause that same harm as a means of bringing about that same good result (see, McIntyre, 2019 for an overview). A central example that illustrates this principle is the now widely discussed "trolley problem", introduced by Philippa Foot (1967), and further popularized by Judith Jarvis Thomson (1975). In this scenario, one imagines a run-away trolley that is headed toward a group of five workers who will be killed if nothing is done. In one half of the problem, you are asked to imagine that there is a lever you can pull to divert the trolley onto a sidetrack. Unfortunately, however, a single worker is on that sidetrack and will be killed if the train is diverted. One then compares the moral permissibility of sacrificing the one to save the five in this case to the permissibility in a second case. In this other half the problem, you are asked to imagine that there is a single person who can be sacrificed by pushing them onto the tracks, which will stop the train and prevent it from killing the other five workers. This widely replicated pattern illustrates that there is an intuitive difference between the two cases, with it being more acceptable to sacrifice the man in the first case vs. the second. The question is how to explain this intuitive difference.

This philosophical distinction between the permissibility of causing harm as a means vs. a side-effect has come to play a central role in empirical moral psychology, initiated largely by the work of Greene and colleagues (Greene 2001, Greene & Cushman 2009, Cushman 2013). At this point, it has not only been argued to support "dual-systems" accounts of moral decision-making (Greene 2008, Cushman 2013), but perhaps more importantly, it has also been shown to emerge early in human development (Levine 2016), to be shared widely across cultures (Barrett 2021), and to have a central role in humans' "moral grammar" (Mikhail 2006). In short, the doctrine of double effect has been pursued as one of the central aspects of moral judgment across philosophy (e.g., Foot 1967), psychology (e.g., Greene 2001; Cushman 2013), and cognitive science (e.g., Klieman-Weiner and Halpern 2018). The importance of this doctrine can also be seen by considering its real-world applications in life-and-death decisions involving self-driving cars (Nyholm 2016), health care (Walker 1991), or resource allocation in disaster scenarios (Shea 2010).

Intriguingly, Judith Jarvis Thomson, who originally introduced the now standard form the trolley problem (Thomson 1976), returned to this dilemma relatively recently and revised her original view (Thomson 2008). The critical new issue that Thomson raised was how to fit our judgments in the original two cases with a third new case in which there are three options, (i) do nothing and let the five workers die, (ii) turn the trolley onto the right sidetrack killing one worker, and (iii) turn the trolley onto the left sidetrack, killing oneself (Thomson 2008). While we will set aside the details of Thomson's argument, a key part of the intuition she relies on is that it would be impermissible to choose (ii) over (iii).

What explains the difference between option (ii) and (iii)? Intuitively, one important difference, which has thus far gone unexplored in the empirical literature on the doctrine of double effect, is that when sacrificing oneself, one consents to being harmed. This idea sparked the question we pursue in the paper: How does consent change moral judgments involving the doctrine of double effect? More specifically, we investigate (1) whether harming someone as a means may become morally acceptable when that person consents to being used as a means, and (2) whether the distinction between harming as a means or side-effect persists in cases of consent to being harmed.

Before turning to the studies we conducted, it is worth noting that consent generally plays an important role in both moral judgment (Sommers 2019) and in real-world moral decisions (Shea, 2010; Walker, 1991). The medical field, for example, has sparked discussions about consent, especially in situations where a life is at risk. With end-of-life treatments, consent may reframe ending a patient's life as simply an example of good palliative care (Allmark 2010). Despite Thomson's opinion-change, there has not been work on moral decision-making when the consent is given by someone that is not oneself. We pursue this question here.

The Present Studies

We collected a set of scenarios used in previous research to demonstrate the means vs. side effect distinction in judgments of moral acceptability (Cushman et al. 2006; Paxman, Ungar, and Greene 2012; Cushman & Greene 2012; Greene 2001). For Study 1, we modified these scenarios, allowing us to introduce a 'consensual means' case that minimally differed from the original non-consensual means cases except that the person consented to being harmed as a means of bringing about some greater good. We predicted both that consent is a factor when judging the acceptability of a harmful action, and more importantly, that an agent's consenting to being harmed as a means will reduce or eliminate the difference in moral permissibility (compared to a side effect case). In other words, we predicted that there would be a smaller difference in acceptability between side effect cases and consensual means cases than between side effect cases and the original nonconsensual means cases. For Study 2, we additionally introduced a 'consensual side effect' case, allowing us to ask whether consenting to being harmed reduces or eliminates the distinction between harming as a means vs. side effect. Both studies were pre-registered prior to being conducted.

Study 1

Methods

Participants Amazon Mechanical Turk was used to recruit 156 participants in the United States, for an average of 50 responses per condition per scenario. Participants were excluded if they did not complete the full survey or completed all 11 scenarios in less than 120 seconds (an average of ~10 seconds per scenario in addition to 10 seconds for the instructions). 6 participants were excluded for not completing the full survey, and 2 participants were excluded for completing the survey too quickly. Analyses were conducted with the remaining 148 participants.

Participants and Materials Participants were tasked with completing an 11-scenario Qualtrics survey about moral dilemmas in which they were asked to rate the appropriateness of choosing the death of the individual over the death of a larger group. Scenarios were presented in random order, and each participant was randomly selected into one of three conditions for each scenario: 'means', 'side effect', and 'consensual means.' Participants presented with

a 'means' or 'consensual means' scenario were also asked to respond to the perceived consent of the individual being sacrificed. Since participants were randomly presented with one condition for each individual scenario, each participant was presented with a mix of conditions throughout the survey. Scenarios were sourced from prior work with the criteria that they had both a 'side effect' case and a 'means' case that logically could be adapted for this study to include consent for the 'consensual means' case (Cushman et al. 2006; Paxman, Ungar, and Greene 2012; Cushman & Greene 2012; Greene 2001).

Means case scenarios included moral dilemmas in which individuals were the direct object of an action taken in order to save the larger group. The intent of the participant's action is to directly harm the individual, with the effect being the safety of the larger group. For example, in the means case of the modified rowboat problem illustrated below, the participant must decide the appropriateness of killing a large fisherman by pushing him off a boat and into freezing water in order to keep the boat from sinking.

"You are in a rowboat with a tour guide while sight-seeing on a lake in Alaska. You notice that three fishermen have overturned their boat nearby. They are now in danger of quickly freezing to death in the icy water. You begin to row over to rescue them when you realize that your boat will not hold you, the tour guide, and the three fishermen.

As you and the tour guide pull in the first two, it is obvious that one of them is nearly twice as heavy as either of the other two. If you keep him on board, your boat will sink, and all of the fishermen will die.

Neither you nor the guide can get out because you are rowing, and the guide is performing CPR.

The only way to save two fishermen is to throw the larger fisherman off the boat and rescue the third one. The large fisherman offers to jump off, but his legs have gone numb from the cold and will need help getting out of the boat.

Is it appropriate to throw the large fisherman off the boat to die in order to save the other two?"

Example 1. Sample problem of a means case scenario sourced from prior literature. The bolded portion is included only for the consensual means case but otherwise omitted.

Consensual means case scenarios are modified versions of the means cases, in which individuals explicitly volunteer to be the direct object of an action taken in order to save the larger group. The individual clearly and explicitly consents to an action that would cause their death but is unable to be the agent of that action and must enlist the help of the participant. The consensual means case of the modified rowboat problem involves the fisherman volunteering to jump off the boat but is unable to due to his legs growing numb from the cold. He enlists the agent in the story to push him off the boat.

Side effect case scenarios included moral dilemmas in which individuals were the indirect object of an action taken in order to save the larger group. The hypothetical action taken by the participant would ultimately lead to the death of the unknowing bystander indirectly. For example, in the modified rowboat problem illustrated above, in which the participant must decide the appropriateness of rowing away from a fisherman in freezing water due to a weight limit on the boat. The act of rowing away directly saves the boat's passengers but causes the certain death of the fisherman.

Measures Each participant was asked to rate the appropriateness of the action that would result in the individual's death in each scenario on a 0-100 sliding scale, with 0 being less appropriate and 100 being more appropriate. For the means and consensual means cases, participants were also asked to rate the extent to which they felt the individual being sacrificed consented to the participant's action, with 0 meaning 'did not consent' and 100 being 'fully consented." The means case consent measure acts as a baseline comparison for the consensual means case in order to determine whether consent was being effectively manipulated.

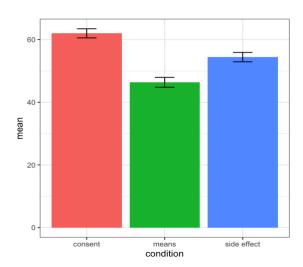
Results

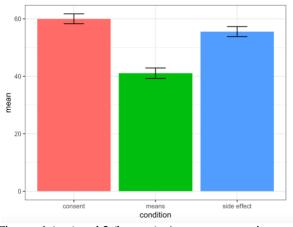
Consent Ratings We first asked whether we succeeded in manipulating perceived consent. We compared a series of linear mixed-effects models predicting participants' consent ratings for the two means cases either with or without condition as a fixed factor. This comparison revealed that our manipulation successfully manipulated consent, $X^2(1) = 224.69$, p < .001.

Appropriateness ratings. We next analyzed participants appropriateness ratings in a similar manner by comparing linear mixed-effects models predicting participants' appropriateness ratings for all three conditions either with or without condition as fixed factor. This revealed participants' appropriateness judgments were significantly affected by condition, $X^{2}(1) = 19.708$, p < .001. We then used the emmeans package to conduct pairwise comparisons between the conditions. We found that participants rated agents' actions as significantly more appropriate in the consensual means condition than in the non-consensual means condition, t(9.97) = 6.536, p < .001, and surprisingly as more appropriate in the consensual means condition than in the side effect condition, t(10.07) = 3.652, p = .011. However, we also found that we only marginally replicated the prior distinction between the means and side effect conditions, t(9.98) = -2.643, p = .059.

Posthoc, we decided to reconduct our analyses only on the subset of 8 scenarios in which the action in the side effect case was judged to be numerically more appropriate than the action in the means case. We once again found that participants' appropriateness judgments were significantly affected by condition, $X^2(1) = 15.082$, p < .001, and again conducted pairwise comparisons using the emmeans package between the three conditions. We found that within this

subset of the data, we replicated the prior finding between the means and side effect conditions, t(6.92) = -3.653, p = .020, and once again found that participants rated agents' actions as significantly more appropriate in the consensual means condition than in the non-consensual means condition, t(6.96) = 6.235, p = .001. We now found that participants only judged actions in the consensual means conditions as marginally more appropriate than actions in the side effect conditions, t(7.00) = 2.499, p = .092.





Figures 1 (top) and 2 (bottom): Average appropriateness ratings for each condition. Figure 1 includes the measures from all 11 scenarios; Figure 2 only includes the 8 scenarios that numerically replicated the means vs. side effect distinction for the non-consensual cases. Error bars depict +/- 1 SEM.

Relationship between consent and appropriateness. Next, we investigated the relationship between participants' perceived level of consent and their appropriateness judgments in the two means cases across all 11 scenarios. We again compared a pair of linear mixed-effects models that allowed us to ask whether, controlling for condition, participants' consent ratings significantly predicted their appropriateness ratings. We found that they did, $X^2(1)$

=133.67, p < .001, see Fig 3. Moreover, 10 of the 11 scenarios showed significant positive correlations, with r values (rounded to 4 significant digits) ranging from 0.284 to 0.552 and p-values ranging from .005 to <.0001. Scenario 9 was the only scenario that was not statistically significant, yielding a correlational value of r = 0.1135, p = 0.2611.

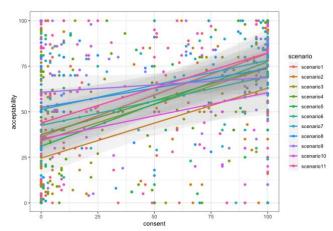


Figure 3: Depiction of the relationship between perceived consent and moral appropriateness in the means conditions for all 11 scenarios.

Study 2

Methods

Participants Amazon Mechanical Turk was used to recruit 207 participants in the United States, for an average of 50 responses per condition per scenario. Participants were excluded if they did not complete the full survey or completed all 7 scenarios in less than 80 seconds (an average of ~10 seconds per scenario in addition to 10 seconds for the instructions). 6 participants were excluded for not completing the full survey. Analyses were conducted with the remaining 201 participants.

Procedures and Materials Similar to study 1, participants were tasked with completing a 7-scenario survey under the three conditions from the aforementioned study (means, side effect, consensual means), with the addition of a consensual side effect case. Scenarios were sourced from Study 1. Scenarios were chosen based on whether they numerically demonstrated the side effect versus means distinction found in prior studies. Additionally, scenarios in which a consensual side effect case could not be minimally created from the original side effect case while still logically making sense were excluded.

Consensual side effect case scenarios are modified versions of the side effect case, in which individuals explicitly volunteer to be the indirect object of an action taken in order to save the larger group. The consensual side effect

case of the modified rowboat problem involves the fisherman shouting at the boat to row away from him as he remains submerged in freezing cold waters.

"You are in a rowboat with a tour guide while sight-seeing on a lake in Alaska. You notice that three fishermen have overturned their boat nearby. They are now in danger of quickly freezing to death in the icy water. You begin to row over to rescue them when you realize that your boat will not hold you, the tour guide, and the three fishermen.

As you and the tour guide pull in the first two, it is obvious that **they are too heavy. If you bring the third fisherman on board**, your boat will sink, and all of the fishermen will die.

Neither you nor the guide can get out because you are rowing, and the guide is performing CPR.

The only way to save two fishermen is to **row quickly** away from the third one, leaving him to die.

The third fisherman, who is still in the water, notices how close the boat is to sinking and shouts at you to just go and leave him.

Is it appropriate to **leave the one fisherman behind to die** in order to save the other two?"

Example 2. This is an example of a consensual side effect/side effect case. The bolded portions are the modifications to the means case scenario that indicate that this is a side effect scenario. The italicized portion is only included in the consensual side effect case and omitted otherwise.

Measures Similar to Study 1, appropriateness and consent are rated 0-100 on sliding scales. All conditions in all scenarios will include the perceived consent measurement so both consensual side effect and consensual means cases have baseline comparisons for perceived consent in the side effect and means cases, respectively.

Results

Consent Ratings We again first asked whether we succeeded in manipulating perceived consent. As before, we compared a series of linear mixed-effects models predicting participants' consent ratings with or without 'consent' condition as a fixed factor (now always including the means-condition as a separate fixed factor). This model comparison revealed that our manipulation again successfully altered perceived consent, $X^2(1) = 36.82$, p < .001.

Appropriateness ratings. We next analyzed participants appropriateness ratings by comparing linear mixed-effects models predicting participants' appropriateness ratings using consent condition, means condition, and their interaction as fixed factors. This revealed a main effect of whether the action harmed someone as a means or a side-effect, $X^2(1) = 11.522$, p < .001, a main effect of whether person consented to being harmed, $X^2(1) = 17.24$, p < .001, and critically a small but significant interaction between the two, $X^2(1) = 10.001$

4.378, p = .036. We then again used the emmeans package to conduct pairwise comparisons between the conditions and decompose the interaction. We found that when the person being harmed consented to being harmed, there was a smaller and non-significant difference between the means and side effect cases (t(8.31) = -2.816, p = .084), while there was a clearly significant effect when the person being harmed did not consent (t(8.44) = -4.459, p = .008).

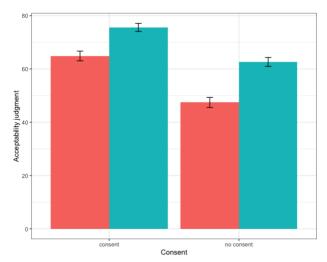


Figure 4: Average appropriateness judgment when the person harmed consented (left bars) or did not consent (right bars) for both harm as a means (red bars) and side effect (blue bars). Error bars depict +/- 1 SEM.

Relationship between consent and appropriateness. Finally, we again investigated the relationship between participants' perceived level of consent and their appropriateness judgments in all 4 conditions means cases across all 7 scenarios. We again compared a pair of linear mixed-effects models that allowed us to ask whether, controlling for consent and means condition, participants' consent ratings significantly predicted their appropriateness ratings. We found that they did, $X^2(1) = 120.79$, p < .001, see Fig 5. Moreover, all 7 of the scenarios showed significant positive correlations, with r values ranging from 0.151 to 0.501 and p-values ranging from .034 to <.001.

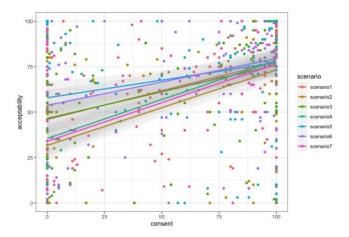


Figure 5: Depiction of the relationship between perceived consent and appropriateness for all 7 scenarios.

Discussion

Across two studies, we asked whether (1) harming someone as a means becomes morally acceptable when that person consents to being used as a means, and (2) whether the distinction between harming as a means vs. side-effect persists even when the person being harmed consents. The results of our studies provide relatively clear answers to these questions. In Study 1, we found that harming someone as a means not only becomes morally acceptable overall but that it becomes roughly as morally acceptable as harming someone merely as a side effect. In Study 2, we replicated this effect of consent, and additionally found that the meansside effect distinction is significantly reduced, though not entirely eliminated, when agents consent to being harmed. Across both studies, we also found that perceived consent was highly predictive of appropriateness ratings, even controlling for condition.

It is worth emphasizing that while these findings demonstrate significant variation in the size of the impact of the doctrine of double effect when it interacts with consent, care should be taken to not overstate this finding because we also find that the DDE consistently exhibits *some* effect across variations in consent. This finding is broadly in line with recent work on variation in moral judgment across cultures (Barrett et al., 2016), which similarly finds that while there is variation in moral judgment, there are also broadly consistent patterns that persist throughout that variation.

Instead, we think these studies demonstrate the importance and potential for continuing research on the role of consent more generally in moral judgment. While it is widely known from prior literature that the presence of consent makes harm seem more permissible, there remain cases in which consent cannot be given in the moment or is given prior to the situation that requires it. For example, future work may be inspired by the real-life events and procedures, particularly within the medical field. DNR orders, for example, involve consent in the complicated calculus medical providers engage in when they weigh the potential good of fixing mistakes or saving a patient's life against the possibility that

they may not succeed and will leave the patient in a worse state than before (Walker 1991).

We hope that future work will take up these questions and continue to examine the role of consent in both the doctrine of double effect and in moral psychology more generally.

References

- Allmark, P., Cobb, M., Liddle, B. J., & Tod, A. M. (2010). Is the doctrine of double effect irrelevant in end-of-life decision making? Nursing Philosophy, 11(3), 170–177. https://doi.org/10.1111/j.1466-769X.2009.00430.x
- Barrett, H. C., & Saxe, R. R. (2021). Are some cultures more mind-minded in their moral judgements than others? Philosophical Transactions of the Royal Society B: Biological Sciences, 376(1838), 20200288. https://doi.org/10.1098/rstb.2020.0288
- Castellano, G. (2007). The Criminalization of Treating End of Life Patients with Risky Pain Medication and the Role of the Extreme Emergency Situation. Fordham Law Review, 76(1), 203.
- Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. Personality and Social Psychology Review, 17(3), 273–292. https://doi.org/10.1177/1088868313495594
- Cushman, F., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. Social Neuroscience, 7(3), 269–279. https://doi.org/10.1080/17470919.2011.614000
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. Psychological Science, 17(12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x
- Di Nucci, E. (2013). Self-Sacrifice and the Trolley Problem. Philosophical Psychology, 26(5), 662–672. https://doi.org/10.1080/09515089.2012.674664
- Everett, Jim A. C. and Pizarro, David A. and Crockett, M. J. (2016) Inference of trustworthiness from intuitive moral judgments. Journal of Experimental Psychology: General, 145 (6). 772-787. https://doi.org/10.1037/xge0000165
- Fink, S. (2009, August 25). The Deadly Choices at Memorial. The New York Times. https://www.nytimes.com/2009/08/30/magazine/30doctors.html
- Foot, P. (1967). The problem of abortion and the doctrine of the double effect. Oxford Review, 5.
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. Cognition, 111(3), 364–371. https://doi.org/10.1016/j.cognition.2009.02.001
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. Cognition, 107(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of

- Emotional Engagement in Moral Judgment. Science. https://doi.org/10.1126/science.1062872
- Halpern, J., & Kleiman-Weiner, M. (2018). Towards Formal Definitions of Blameworthiness, Intention, and Moral Responsibility. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1), Article 1. https://ojs.aaai.org/index.php/AAAI/article/view/11557
- Huebner, B., & Hauser, M. D. (2011a). Moral judgments about altruistic self-sacrifice: When philosophical and folk intuitions clash. Philosophical Psychology, 24(1), 73–94. https://doi.org/10.1080/09515089.2010.534447
- Huebner, B., & Hauser, M. D. (2011b). Moral judgments about altruistic self-sacrifice: When philosophical and folk intuitions clash. Philosophical Psychology, 24(1), 73–94. https://doi.org/10.1080/09515089.2010.534447
- Kamm, F. M. (1999). Physician-Assisted Suicide, the Doctrine of Double Effect, and the Ground of Value. Ethics, 109(3), 586–605. https://doi.org/10.1086/233923
- Kawai, N., Kubo, K., & Kubo-Kawai, N. (2014). "Granny dumping": Acceptability of sacrificing the elderly in a simulated moral dilemma. Japanese Psychological Research, 56(3), 254–262. https://doi.org/10.1111/jpr.12049
- McIntyre, A. (2019). Doctrine of Double Effect. The Stanford Encyclopedia of Philosophy, Spring 2019. https://plato.stanford.edu/archives/spr2019/entries/double-effect/
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. Trends in Cognitive Sciences, 11(4), 143–152. https://doi.org/10.1016/j.tics.2006.12.007
- Nyholm, S., & Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem? Ethical Theory and Moral Practice, 19(5), 1275–1289. https://doi.org/10.1007/s10677-016-9745-2
- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. Cognitive Science, 36(1), 163–177. https://doi.org/10.1111/j.1551-6709.2011.01210.x
- Shea, F. (2010). Hurricane Katrina and the Legal and Bioethical Implications of Involuntary Euthanasia as a Component of Disaster Management in Extreme Emergency Situations. Annals of Health Law and Life Sciences, 19(1), 133.
- Shea, F. K. (2010). Hurricane Katrina and the Legal and Bioethical Implications of Involuntary Euthanasia as a Component of Disaster Management in Extreme Emergency Situations. Annals of Health Law, 19, 8.
- Sommers, R. (2019). Commonsense Consent. Yale Law Journal, 129(8), 2232–2325.
- Thomson, J. J. & The Hegeler Institute. (1976). Killing, letting die, and the trolley problem: Monist, 59(2), 204–217. https://doi.org/10.5840/monist197659224
- Thomson, J. J. (2008). Turning the Trolley. Philosophy & Public Affairs, 36(4), 359–374. https://doi.org/10.1111/j.1088-4963.2008.00144.x

Walker, R. M. (1991). DNR in the OR: Resuscitation as an Operative Risk. JAMA, 266(17), 2407–2412. https://doi.org/10.1001/jama.1991.03470170095031