

Meta-Reasoning in Making Moral Decisions under Normative Uncertainty

TOMASZ ŻURADZKI

Institute of Philosophy, Jagiellonian University, Poland

t.zuradzki@uj.edu.pl

In my paper I analyze recent discussions about making moral decisions under normative uncertainty. I discuss whether this kind of uncertainty should have practical consequences for decisions and whether there are reliable methods of reasoning that deal with the possibility that we are wrong about some moral issues. I defend a limited use of the decision theory model of reasoning in cases of normative uncertainty.

KEYWORDS: bioethics, decision theory, metaethics, moral uncertainty, the ethics of war

1. INTRODUCTION

The simple fact is that even when we feel quite certain about moral issues, we are susceptible to mistakes. For example, we have to act in the face of uncertainty about the facts, the consequences of our decisions, the identity of people involved, people's preferences, moral doctrines, specific moral duties, or the ontological status of some entities (belonging to some ontological class usually has serious implications for moral status). I want to analyze whether these kinds of uncertainties should have practical consequences for actions and whether there are reliable methods of reasoning that deal with the possibility that we understand some crucial moral issues wrong.

The most promising approach is to try to extend the decision theory model of reasoning to encompass normative uncertainty (Lockhart, 2000). But this model, when used to guide our action in moral terms, is highly controversial. It assumes that in the face of normative risk or uncertainty we are rational if and only if we maximize expected value (whatever it is). In this case two things would determine what we ought to do under normative uncertainty: 1) the probabilities

assigned to the various normative views; 2) the differences in values between the available actions, according to each of those views. This approach – if successful – would have some interesting applications both in metaethics (e.g. rejecting nihilism, see: Ross, 2006) and applied ethics in particular the ethics of war and bioethics (permissible killing people or animals, abortion, embryo research, see: Guerrero, 2007; Moller, 2011; Friberg-Fernros, 2014; Żuradzki, 2012, 2014). Moreover, the argument (from moral or normative uncertainty) is also used in other contexts (Henning, 2015).

The main problem that the supporters of this approach have to deal with is the question of intertheoretic comparisons of value. Normally, when we use expected utility calculus we use a common scale by which it is possible to measure the values attached to different outcomes (this is one reason why so many examples refer to money). But there doesn't seem to be any way of making this kind of intertheoretic comparison of moral values between different theories or doctrines. Moreover, recently a few philosophers presented other arguments against the attempts to extend expected value maximization style reasoning to encompass moral uncertainty (MacAskill, 2013; Weatherson, 2014; Harman, 2015; Nissan-Rozen, 2015, Hedden, forthcoming).

2. MANY FACES OF UNCERTAINTY

There are many ways in which you can be uncertain about the morally important aspects of your action. For example, you could be certain that some action can harm some people (or other beings with the moral significance), but uncertain about the identity and the number of people involved.

A combatant who fires indiscriminately from his gun at a place inhabited by noncombatants is in this position. He knows that there is a serious risk that some innocent people could be fatally shot, but he knows neither their identity nor number. Alternatively, you can know that there is a serious risk that some action will harm a person and you can know who is this person. Someone who plays Russian roulette with a prisoner of war is in this position.

Finally, you can be sure that some action will harm some identified beings, but not sure about its ontological or moral status. The real-life examples of beings with uncertain ontological or moral status (at least from the perspective of some people) include fetus, human embryos (in particular at the very early stages of development) or some products (real or only possible) of genetic engineering. Some philosophers have also argued that it is the case of at least some

animals. Analogically to the two above examples, one could be tempted to say that someone who aborts fetus, destroys human embryos (during in vitro procedures or during scientific research) or kills animal is in a similar position as soldiers from the two above examples.

In all these cases there is a risk that something that has moral status will be harmed. So the popular argument says that in all these three cases there are strong reasons for agents not to act: not fire indiscriminately at area inhabited by noncombatants, not play Russian roulette with a prisoner of war, not harm human embryos or animals. And this popular argument adds that it is also a reason to condemn such actions no matter what their results are. It is wrong – says this argument – to impose a risk to someone who is not liable to be exposed, even if a potential victim is actually not harmed (review article about the problem of risk imposition: Hayenhjelm & Wolff, 2012). This argument – presented very often by Catholic preachers or scholars (but also, surprisingly, by defenders of animals rights) – treats all three above cases in a very similar way and is used to argue against the permissibility of abortion or destruction of early human embryos. It is usually presented in a form an analogy to hunting:

Example 1: Deer hunting

If I am hunting with a rifle, and I see something move in the trees but am unsure whether it is a deer or a person, I am obliged not to shoot until I establish that it is in fact a deer: better safe than sorry (Shaw, 2008, p. 219).

Catholic preachers or scholars (and vegetarians) argue that when someone is unsure whether some being has a full moral status or not, he should be obliged not to kill it. Since it is impossible to resolve empirically whether or not the target has full moral status, obligation not to kill is not time-limited. According to this view (see for example: Friberg-Fernros, 2014) the same argument can be applied to human embryos (in particular in early stages): since there are reasonable doubts about their personal status, morality requires that human embryos from conception be treated as persons.

3. FACTUAL UNCERTAINTY AND THE ETHICS OF WAR

In this chapter I will use some examples from the ethics of war to demonstrate the difference between cases of factual and normative uncertainty. In a war context the question of risk imposition is discussed usually in this kind of context:

Example 2: A security checkpoint

Imagine that you are a soldier, ordered to protect a military or diplomatic convoy as it passes through hostile territory, and you see a car stopped by the side of the road ahead. Or imagine that you are stationed at a security checkpoint and a car approaches despite signs and warnings directing it to stop. The occupants of the car may be civilians, but they also may be irregular forces waiting to attack (Haque, 2014, p. 65).

So the question in this case is how certain should be a soldier that people at the front of him are combatant, rather than noncombatants, before using deadly force? Surely, soldiers are in a different position than hunters, because the stake is much bigger: they risk their lives and they risk the case for which they fight. In a hunter case there is hardly anything valuable in killing a deer (except from hunter's pleasure, since I assume that it is not necessary hunting). In this sense a soldier's case is more similar to some cases of abortion or embryo research, where the stake also can be quite high (the well-being or health of woman; development of science during embryo research).

The obligation not to kill civilians is well-established in international law and international theory. For example Protocol Additional to the Geneva Conventions says:

In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives (Protocol, 1977, art. 48).

So firstly, all participants in conflicts must determine whether an aim is legitimate or illegitimate target: they must distinguish between combatants and noncombatants (I will assume that combatants are people who are directly involved in hostilities, some of them can be civilians). And secondly, it is permissible to target only combatants (or other military objectives). Of course this description is highly idealized, and I we have seen in *A security checkpoint* case very often it unclear whether the target of operation are combatants or not. Surprisingly, international law proposes a very similar rule to this one that was proposed in *Deer hunting* case:

In case of doubt whether a person is a civilian, that person shall be considered to be a civilian (Protocol, 1977, supra note 9).

Those who plan or decide upon an attack shall (...) do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives (Protocol, 1977, supra note 5).

This point of the Protocol does not specify any the level of care with which a soldier must try to distinguish combatants from noncombatants. Neither this point specify how much effort should be put in verifying (soldiers should only do “everything feasible”). In its literal interpretations it means that if there is any doubt about the nature of a target proceeding is prohibited. It would mean for example that in cases like *A security checkpoint* when there is any doubt whether in an approaching car are noncombatants, soldiers are not allowed to use deadly force.

In many conflict situation this level of civilian protection would be too restrictive and some Western countries entered reservation about this provision (for example the UK states that it “applies only in cases of substantial doubt” - Declaration 2002). Haque noticed that despite literate meaning of this regulation in the literature or commentaries the most common approach to this problem is the balancing approach according to which “both the required level of certainty and the required level of risk vary with the balance of military and humanitarian considerations” (Haque, 2012, p. 63). This approach is visible in Walzer’s *Just and Unjust Wars* when he writes about combatants’ obligations:

The degree of risk that is permissible is going to vary with nature of the target, the urgency of the moment, the available technology, and so on (Walzer, 2000, p. 156).

It is also proposed by commentators who underline that combatants should “balance” possible benefits and risk to civilians:

The reasonable care rule is disquieting. It vests belligerents with considerable discretion in multifaceted balancing and legitimizes even large-scale injury to innocent civilians under certain circumstances (Waxman, 2012, p. 1393).

So the most common approach among scholars is quite contradictory to the literal meaning of the Protocol: now one expects that soldiers can act only if he is 100 percent certain in all cases. In some circumstances soldiers can attack even in the face of substantial doubts about the status of target (when the military stakes are very high), and in some circumstances soldiers must abstain from attack in the face of even very slight doubts.

It can be visible in the next hypothetical example, which is slightly modified version of the Example no. 1 *A security checkpoint*:

Example 3: A security checkpoint (modified)

Imagine that you are a soldier, ordered to protect a military or diplomatic convoy [consisting of: a) 5 people; b) 20 people] as it passes through hostile territory, and you see a car stopped by the side of the road ahead. Or imagine that you are stationed at a security checkpoint [consisting of: a) 5 soldiers; b) 20 soldiers] and a car approaches despite signs and warnings directing it to stop. The occupants of the car may be civilians, but they also may be irregular forces waiting to attack.

In this case we modified number of soldiers endangered: in the a) cases there are only 5 endangered soldiers, in the b) cases there 20 endangered soldiers. I assume that if the occupants of the car are irregular forces they will want to kill all soldiers. According to the balancing approach the way in which soldiers should precede depends on the number of soldiers endangered: they are permissible to act even if they have more serious doubts in the b-type cases than in the a-type cases.

There are serious objections to the balancing approach in case of the ethics of war (Haque, 2012), but I am not going to evaluate them. Instead, I want to demonstrate that this kind of balancing approach is even more difficult to use in the cases of moral uncertainty.

4. MORAL UNCERTAINTY AND BALANCING APPROACH

Let me start this part with a typical example discussed in the literature.

Example 4: Meat eating

Suppose, for example, that an agent is uncertain between two views about the morality of eating meat. In one view, eating meat is tantamount to murder; it is much, much worse, then, to eat meat than to abstain from it. In another view, it is ever so slightly better to eat meat than to abstain – better, perhaps,

for reasons of health or pleasure. In the most plausible views of rationality under moral uncertainty, it is rational to avoid eating meat, even if one's belief in the second view is slightly higher (Sepielli, 2013, p. 581).

At first sight it seems that this example is very similar to the case of *Deer hunting*: on the one hand we can act in a risky way – we can have a small benefit (a deer is ours; we eat a nice meal), but there is a serious possibility that we do something very wrong: we kill a person; or we kill a being that have full moral status – the same or almost the same as an adult person. So it may seem that rationality requires to do the action with “the highest expected moral value”. What does it mean? “An action's expected moral value is the probability-weighted sum of its moral values according to the various moral views or theories” (Sepielli, 2013, p. 581). Moreover, it seems that any plausible theory of rational decision making under uncertainty will care about moral stakes of decisions according to different moral theories. So, below I demonstrate that the expected moral value solution is only one of many possible solutions that were proposed recently.

It is worth noting that other examples related with decision-making under normative uncertainty (abortion, destroying early human embryos) are more complicated than meat eating because there are important values not on one side, but on both (so they are more similar to the security checkpoint cases). Abortions are usually defended because of some important values at stake (well-being or health of a woman, the low quality of life of possible child). The same is with destroying of early human embryos either because of reproductive purpose (in vitro fertilization) or because of scientific research. Below we will see the importance of this difference.

4.1 My Favorite Theory and its problems

Probably the most obvious proposition how to act under normative uncertainty is My Favorite Theory approach. It says that “a morally conscientious agent chooses an option that is permitted by the most credible moral theory” (Gustafsson & Torpman, 2014, p. 159). Even if you have doubts about meat eating, you are permitted to eat it if you believe that theory according to which animals have not any important moral status is more reliable than any other theory about animal status.

Although this approach looks very intuitive, there are interesting counter-examples. Consider the following case (adopted from (Gustafsson & Torpman, 2014) in which you can either choose action A or action B, and your credences are divided between two moral

doctrines MD1 and MD2. You think that MD1 is slightly more reliable than MD2 (I introduce some numbers for convenience, for example that your credence in MD1 is 0,6 and your credence in MD2 is 0,4, but of course in real life this kind of precision is not necessary).

	A	B
MD1 (0,6)	Slightly wrong	Merely OK
MD2 (0,4)	Saintly	Morally terrible

Table 1

The descriptions of the results show the moral evaluation of possible outcomes (wrong, merely OK...). In this kind of cases it seems that despite the fact that an agent thinks that the moral theory MD1 is more reliable than MD2, she should act as if the theory MD2 were correct or right moral doctrine. It would mean that she should prefer action A over action B – against My Favorite Theory. Why? Because – analogically to our previous case of *Meat eating* – an agent may prefer not risking any serious moral wrongdoing. If she decided on B, which seems OK according to the MD1, the favorite theory for an agent, she risk a serious wrongdoing if she were not right about moral theories and in fact MD2 were the right doctrine.

As in our previous cases related to the ethics of war, this case also assumes the balancing account: the possibility of making comparisons and weighing different values on one scale. In recent literature there are two approaches to this problem: non-comparativism and comparativism.

This first position (Sepielli, 2013; Nissan-Rizan, 2014) assumes that there is no analogy between factual and normative uncertainties and it is impossible to make intertheoretic comparisons of values. It means that tables like Table 1 have no sense, because the moral evaluation of results (wrong, merely OK...) is different for MD1 and MD2. So it is meaningless to say that rationality requires that we avoid doing something morally terrible (according to MD1), since there is no common “currency” for both moral theories. This common currency is a necessary requirement

In the previous cases related to the ethics of war we compared the values of soldiers’ lives, the importance of military target, and the value of noncombatants lives. Even if it is very hard to compare how much value “one civilian” versus some military target, the balancing account tries to do it. In cases of normative uncertainty this kind of comparisons are even harder: it seems that there is no way for making

this kind of intertheoretic comparisons of moral values between different theories or doctrines.

4.2 Expected value approach and dominance principle

Comparativism in its strong form (Lockhart, 2000; Sepielli, 2009) is now an unpopular position. The main assumption of this view was to calculate the expected moral values of the available actions, relative to possible axiologies, and summing up those expected moral values, weighted by the degree of belief that the corresponding axiology is correct. There were two main propositions how to do it: the principle of equity among moral theories (Lockhart, 2000), the reactive-attitude approach (Sepielli, 2009). Both of them seem to be obviously mistaken (Sepielli, 2013).

Comparativism in its weak form seems to be much better proposition. It has weak form because it does not require any calculation of expected moral values of all available actions. It means that it can be applied only to very specific kinds of situations in which an agent’s credences are not divided between two different moral doctrines, but between only one moral doctrine and some doctrine (or doctrines) that does not give any moral reasons. Its conclusion says that if some theories in which you have credence give you subjective reason to choose action A over action B, and no theories in which you have credence give you subjective reason to choose action B over action A, then you should (because of the requirements of rationality) choose A over B (Ross, 2006).

Let me introduce another example in which this type of reasoning should work perfectly well (this is a modified version of the example discussed by Ross, 2006). Suppose, for example, that John must decide whether to kill some being (that could have important moral status) or not. An agent strongly believes (it is doctrine no. 1 – D1) that it is highly probable that from the moral point of view it does not matter if he kills this type of being or not, but he is not absolutely certain of his normative views. Let me assume that his degree of credence is 0.99. This means that he thinks that there is a very small chance that another doctrine is the right one (D2). According to this second doctrine killing this kind of organism is in fact morally terrible (his degree of credence regarding this view is 0.01).

	A	B
D1 (0,99)	Does not matter	Does not matter
D2 (0,01)	Morally right	Morally terrible

Table 2

Accepting My Favorite Theory approach it does not matter what we choose because according to our favorite theory there are no reasons in favor or against both possible options. In this case we could – for example – decide by flipping a coin. But an agent accepts doctrine that says that there is very small chance (0,01) that it would be morally terrible to kill this type of organism and morally right not to kill it. If an agent wants to maximize the expected moral value of his decisions, he should choose A, even though he believes that the probability that killing this type of being is morally wrong is indeed extremely low.

Why this case is different than described in the Table 1? The application of expected moral value approach in this type of reasoning seems to be correct, because here there is no problem with intertheoretic comparisons of values. An agent does not have to compare in this situation any values or disvalues between different moral doctrines or views on the moral status of this living organism, since one of views says that everything he does in the situation is morally neutral. So it seems that in these types of cases, the ANU would indeed give a reason to prefer the safer option, only if the probability that this option is morally correct is greater than zero.

5. CONCLUSION

In this paper I sketched the problem of meta-reasoning in making moral decisions under normative uncertainty. I found that there is a promising type of meta-reasoning proposed by Ross (2006) that could be applied to some cases of normative uncertainty. However, it is worth noting that this kind of meta-reasoning has been recently extensively criticized. In this paper – because the lack of space – I have not considered some important arguments against meta-reasoning. In recent literature there are at least five such critiques. The first refers to the problem of “the infectiousness of nihilism” (MacAskill, 2013). This argument says that if we have nonzero credence in nihilism it is impossible to use expected value reasoning (also in this weak sense of comparativism) in the situations of normative uncertainty, because the expected value of all options is undefined. The second refers to the problem with accessibility. Even if descriptive facts may often be inaccessible to agents, we could assume that normative facts are a priori, then there is a sense in which any agent is in a position to know the moral truth (Hedden, forthcoming). In this case there would be no such situations as normative uncertainty. Thirdly, some authors say that although non-culpable factual ignorance is an excusing factor, it is not the case with normative ignorance which does not exculpate (Harman, 2014). The

conclusion would be similar to the first critique: there is no such state as normative uncertainty. Fourthly, there is an argument from action-guiding and fetishism argument. It says that morally good people care non-derivatively about other people, their well-being and the like. Meta-reasoning would force agents to care not about people but about one thing: doing what they believe to be right, where this is read *de dicto* and not *de re* (Weatherson, 2014). If this critique is right probably the best theory under normative uncertainty would be My Favorite Theory account. And fifthly, as one author has just noticed there is an additional problem related with risk attitudes of agents, who are normatively uncertain. In many situations they could assign positive credence to several theories with different attitudes toward risk. In this kind of cases the intertheoretical comparisons of moral value would be meaningless in the same way that sentences like 'thunders are louder than honey is sweet' are meaningless. (Nissan, 2015, p. 358).

ACKNOWLEDGEMENTS: This work was supported by National Science Centre (NCN) grant SONATA number UMO-2011/03/D/HS5/01152.

REFERENCES

- Declaration, (2002), United Kingdom Reservation / Declaration of July 2, 2002 to Protocol I, *supra* note 9, available at: <http://www.icrc.org/ihl.nsf/NORM/0A9E03F0F2EE757CC1256402003FB6D2?OpenDocument>
- Friberg-Fernros, H. (2014). Taking precautionary concerns seriously: A defense of a misused anti-abortion argument. *Journal of Medicine and Philosophy*, 39(3), 228-247.
- Guerrero, A. A. (2007). Don't know, don't kill: Moral ignorance, culpability, and caution. *Philosophical Studies*, 136(1), 59-97.
- Gustafsson, J. E., & Torpman O. (2014). In defence of My Favourite Theory. *Pacific Philosophical Quarterly*, 95(2), 159-174.
- Harman, E. (2015). The irrelevance of moral uncertainty. In: R. Shaffer-Landau (Ed.), *Oxford Studies in Metaethics*, vol. 10 (pp. 53-78). Oxford: Oxford University Press.
- Haque, A. A. (2012). Killing in the fog of war. *Southern California Law Review* 86(1), 63-116.
- Hayenhjelm, M., & Wolff, J. (2012). The moral problem of risk impositions: A survey. *European Journal of Philosophy*, 20, 26-51.
- Hedden, B. (forthcoming). Does MITE make right? On decision-making under normative uncertainty. In: R. Shaffer-Landau (Ed.), *Oxford Studies in Metaethics*, vol. 11, Oxford: Oxford University Press.

- Henning, T. (2015). From choice to chance? Saving people, fairness, and lotteries. *Philosophical Review*, 124(2), 169-206.
- Lockhart, T. (2000). *Moral uncertainty and its consequences*. Oxford: Oxford University Press.
- MacAskill, W. (2013). The infectiousness of nihilism. *Ethics*, 123(3), 508-520.
- Moller, D. (2011). Abortion and moral risk. *Philosophy*, 86(3), 425-443.
- Nissan-Rozen, I. (2015). Against moral hedging. *Economics and Philosophy*, 31(3), 349-369.
- Protocol (1977). Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), June 8, 1977, available at: <https://www.icrc.org/ihl/INTRO/470>
- Ross, J. (2006). Rejecting ethical deflationism. *Ethics*, 116(4), 742-768.
- Sepielli, A. (2009). What to do when you don't know what to do. In: R. Shaffer-Landau (Ed.), *Oxford Studies in Metaethics*, vol. 4 (pp. 5-28). Oxford: Oxford University Press.
- Sepielli, A. (2013). Moral uncertainty and the principle of equity among moral theories. *Philosophy and Phenomenological Research*, 86(3), 580-589.
- Shaw, D. M. (2008). Moral qualms, future persons, and embryo research. *Bioethics*, 22(4), 218-223.
- Walzer, M. (2000). *Just and unjust wars. A moral argument with historical illustrations* (3rd edition). London: Basic Books.
- Waxman, M. C. (2008). Detention as targeting: Standards of certainty and detention of suspected terrorists, *Columbia Law Review*, 108(6), 1365-1430
- Weatherston, B. (2014). Running risk morally. *Philosophical Studies*, 167(1), 141-163.
- Żuradzki, T. (2012). Argument z niepewności normatywnej a etyczna ocena badań naukowych wykorzystujących ludzkie embriony. *Diametros*, 32, 131-159.
- Żuradzki, T. (2014). Moral uncertainty in bioethical argumentation: A new understanding of the pro-life view on early human embryos. *Theoretical Medicine and Bioethics*, 35(6), 441-457.