

Computerised Consequentialism to Support Moral Reasoning and Decision Making in Crisis Management

Niklas Humble, Peter Mozelius
Mid Sweden University, Östersund, Sweden
niklas.humble@miun.se
peter.mozelius@miun.se

Abstract: Under normal conditions, humans are good at logic reasoning and taking appropriate decisions, but crisis management is not conducted under normal conditions. In complex and stressful situations, the moral aspects of crisis management could be hard to sort out, with decision makers stuck in moral dilemmas. Research has recommended that the detailed emergency handling manuals should be complemented with decision making support systems. This paper presents and discusses a model for an intelligent decision support system (IDSS), based on the ethical theory of consequentialism.

Consequentialism posits that moral decisions can be calculated by the expected outcomes from presumptive actions. On the other hand, consequentialism has been criticised for deficiencies in the prediction of complex situations. Calculations in the suggested model are inspired by Hooker's Rule-consequentialist theory of morality. A theory with the fundamental idea that potential long-term and short-term outcomes should be weighed against each other. Furthermore, the positive and negative outcomes from presumptive actions should be estimated and weighed against each other.

The suggested model should be flexible enough to be used in both real crisis situations and crisis management exercises. However, before testing the model in real life crisis situations there is a need for thorough evaluations in virtual crisis management exercises. As a complement, these virtual crisis exercises should preferably also include scripted collaboration. The most interesting specialisation of scripted collaboration are the conflict scripts, enabling the idea of scripting exercise conflicts that resemble moral dilemmas in philosophy.

Keywords: Computerised consequentialism, Decision support system, Artificial intelligence, Crisis management, Conflict scripts, Moral dilemmas

1. Introduction

Persons involved in crisis management are partly selected by their proficiency in logic reasoning, and by their decision-making ability, but crisis management is often conducted under stressful conditions. As identified in the study by Weisæth, Knudsen & Tønnessen (2002), stress appears both at the operative level, and at the strategic and political level of crisis management. Their conclusion was that stress had a negative effect on decision making at both these levels (Weisæth, Knudsen & Tønnessen 2002). In complex and stressful situations, the moral reasoning can be hard to sort out, with decision makers stuck in moral dilemmas. Furthermore, without any stress management support the stress can affect the involved persons' mental health and lead to acute stress, burnout, absenteeism, and post-traumatic stress (Guenther, 2012).

Severe crises involve threats to lives, environment and other important goods that require swift actions with instant decision-making that further increase the level of stress (Hofinger, Zinke & Künzer, 2011). Furthermore, decision-making on inadequate basis of information under time pressure with anticipated consequences of wrong decisions can also add more stress (Zinke, Hofinger & Strohschneider, 2010). Trapped in an escalating stress spiral the perceived lack of action just increases the stress level. A recommendation has been to complement the detailed emergency handling manuals with decision making support systems. The aim of this paper is to present and discuss a model for an intelligent decision support system (IDSS), that is based on the ethical theory of consequentialism. Moreover, the paper presents several ethical theories, followed by a comparison and discussion on which of the theories that are computable and suitable to build an IDSS around.

2. Ethical theories

According to Hursthouse and Pettigrove (2016) it is common to divide ethical theories in three main approaches based in their different emphasis on what makes an action morally justified: do they put emphasis on the character, the action, or the consequences? The three main ethical theories are: virtue ethics, deontology, and consequentialism (Hursthouse & Pettigrove, 2016). Another ethical theory that is relevant to the conflicts and debates of the real world, although rejected by many ethicists, is ethical relativism; The theory holds that what is morally right or wrong is dependent on the cultural norms in which the action is performed (Velasquez et al., 1992). It should also be mentioned that although ethical theories are often presented as rival theories, trying to answer the same questions, they do not have to be. Kagan (1992) argues that ethical theories often are concerned with different aspects in ethics and that they are in principle often compatible.

2.1 Virtue ethics

The emphasis on virtues or the moral character of an action is central in virtue ethics. An action is morally justified if the action is benevolent or charitable. Plato and Aristotle are usually considered to be the founders of virtue ethics in the West (Hursthouse & Pettigrove, 2016). The theoretical structure of virtue ethics was first stated clearly by Aristotle (Annas, 2006). According to Aristotle, an act should not be isolated only to itself or to its consequences; these work together in a series of actions and how they affect and develop the character of the one who performs them (Koehn, 1995).

Four virtues are often derived from the works of Plato and Aristotle, the cardinal virtues: prudence, justice, temperance, and fortitude. The cardinal virtues identify which characteristics that constitutes a virtues person. That is, one who is wise, fair, moderate and brave. But these cardinal virtues have been questioned by scholars, for example in their relevance to guide us and our lives in modern society. A productive field within virtue ethics has been to study and identify old and new cardinal virtues. (Hursthouse & Pettigrove, 2016; Keenan, 1995)

2.2 Deontology

The ethical approach of deontology puts emphasis on the word duty in determining what is morally required. The theory focuses more on what actions are permitted, required, and forbidden; rather than what characteristics the person who performs them have (Alexander & Moore, 2020). Deontological ethics are often portrayed as the rival theory of consequentialism regarding guiding moral choices (Alexander & Moore, 2020). A central name in deontology is Immanuel Kant, who stated that there are duties for humans to follow, these duties are universal since they stem from human rationality; Kant formulated the most important duties, those who are absolute, in what is called *categorical imperative* (Misselbrook, 2013; Kant, 1785). For example, "Act only by that maxim by which you can, at the same time, will that it be a universal law" (Misselbrook, 2013).

The divine command theory is also a possible approach within deontological ethics (Alexander & Moore, 2020). There are many versions of divine command theory, but they all rely on some sort of duties or obligations; These obligations are often formulated in the sense that they are commanded by a God (Wierenga, 1983). An important task within the divine command theory discussion is to determine the relationship between the *command* and the *morality*; That is, is something good because it is commanded or is something commanded because it is good? (Murphy, 1998; Wierenga, 1983).

2.3 Consequentialism

As the name suggests, consequentialism is based on the notion that the consequences are what determines if an action is morally justified or not (Sinnott-Armstrong, 2019). An early theory with this approach is Epicureanism which state that what should guide us too good and bad are the sensations of pleasure and pain, since all creatures that are sentient will be drawn and repelled by these sensations; The goal of Epicureanism is therefore to maximise pleasure and minimise pain (Konstan, 2018). Utilitarianism is by many considered a paradigm case within the consequentialist approach (Sinnott-Armstrong, 2019). It is further considered to be one of the most persuasive and influential ethical theories within philosophy (Driver, 2014). Utilitarianism, as epicureanism, holds that we should strive to maximise the good, but are to a greater extent agent neutral and values the qualities and long-term benefits of actions; That is, my happiness is worth as much as any other's and the long-term good of studying is worth more than the short-term pleasure of going to a party (Driver, 2014).

Rule-consequentialism, as regular consequentialism, puts emphasis on the consequences in determining what is morally good or morally wrong; However, as typical consequentialist theories focus on the consequences of specific actions, rule-consequentialism puts its focus on rules that will generally generate good outcomes (Hooker 2015). This allows rule-consequentialism to partly escape the criticism of consequentialism to endorse abominable actions if the expected outcomes are good. A criticism that has been directed towards rule-consequentialism is that it can result in disasters, if people blindly are following rules and do not mind the consequences of the individual actions at all (Hooker 2015). This objection has led to that some rule-consequentialist theories have applied a rule for avoiding disasters, even if following that rule means breaking other rules; This rule comes in play when a true disaster is potential and not when only small gains are to be made (Hooker 2015; Hooker, 2002, p.86).

2.4 Relativism

Relativism is the ethical theory that builds upon the idea that there are no universal norms that can tell the difference between right and wrong. A postulate that leads to that different groups can have contradictory views of a moral problem, and that both can be right. Moreover, relativism can be further divided into the sub-branches of subjective relativism and cultural relativism. Subjective relativism posits that each individual decides what is right for the actual individual, often backed up with the wide-spread expression "What's right for you may not be right for me". (Quinn, 2014, p. 84)

Cultural relativism is when right and wrong are defined by a society's moral guidelines. The idea that one's judgments are based on enculturation, makes it impossible to pass judgments on any other ethical system. As an example, it would be problematic to use cultural relativism in a try to address Adolf Hitler's Nazi morality, or to analyse the ethnic cleansing in Eastern Europe. Another issue is how to resolve the inevitable conflict that arises between different clashing ethical systems? (Simpson, 2001)

3. Computational ethical models

The integration of ethical theory with computer simulations are sometimes referred to as *computational ethics*. This approach puts the computational perspective on ethics and provides a mechanism to test and experiment with ethical principles and theories; Through computer simulations, computational ethics could be used as a tool for enabling quantitative research on ethical theories, for example, exploring large scale consequences of certain actions and finding hidden moral dilemmas (Ruvinsky, 2007). Computational ethics also have the potential to serve as embedded ethics in artificial intelligent systems (Segun, 2021).

An obvious challenge with computational ethics, especially if it is to be used for guiding moral action, is that there are several ethical theories that it could be based on; and there is no consensus on which theory, if any, could be universally applied (Segun, 2021). The ethical theories of Kantianism and utilitarianism have been the main focus of computational ethics, but there are also other examples (Segun, 2021). Deontological theories, such as Kantianism, could be relevant for the basis of computational ethics since it is rule-based and therefore relatively easy to program; However, the use of a rule-based theory has its challenges, such as the need for excessive specification of rules and their priorities (Powers, 2006).

Consequentialist theories, such as utilitarianism, are an appealing choice for computational ethics since it would offer a way to systematise moral judgement in an easy calculative theoretical model; It is also difficult to argue against the end goal of utilitarianism, to make the world a happier place (Grau, 2006). But there are problems with this approach that should produce caution in relation to computational ethics; For example, computational ethics based on utilitarianism could be vulnerable in complex and prior unsolved problems in ethics since the calculations may lead to decisions that are considered immoral by humans (Brundage, 2014). It could therefore be potentially dangerous to put a utilitarian robot in charge of decisions outside of a carefully scoped and narrowed domain (Brundage, 2014). "Since the ends justify the means, the means can get ugly" (Grau, 2006).

There are also examples of combining computational ethics, for example in robots, with virtue ethics (Segun, 2021). Virtue ethics is also used in research to argue for better treatment of social robots, not because they are sentient but because the actions performed say something about the moral characteristics of the one who performs them (Cappuccio, Peeters & McDonald, 2020). A problem with virtue ethics in relation to computational ethics, is that it can be argued to have the same challenges that is brought forwarded by Powers

(2006) in relation to deontological theories; that it requires excessive specification of virtues, instead of rules, and their priorities. The same argument could also be raised against relativism in relation to computational ethics. But in the case of relativism the rules would also have to be specified in relation to each cultural context in which they are to be applied, making the specification even more excessive and complex. This may be a reason for why relativism is not that common as the ethical basis for computational ethics.

4. The suggested computational model

A problem with strict rule-based models that has been highlighted by the just consequentialist James Moor, is if ethical rules are regarded as binding and without exceptions. Kantian ethics with the idea of rules as universal laws does not allow any exceptions. Consequentialism that does not require complete consensus in all situations allows to override rules if the outcomes are extraordinary bad, and if there are human lives at risk. A computational model for crisis management needs rules with exceptions for conflicts and extraordinary consequences. (Quinn, 2014, p. 134-135)

As discussed in previous section, building a computational model for an intelligent decision support system (IDSS) based on virtue ethics, deontology, consequentialism, and relativism would all come with challenges. The problem with building a computational model based on a theory that relies on rules is that it requires excessive specification of rules and their priorities (Powers, 2006). This would be a challenge for all ethical theories discussed in this paper, except consequentialism. A computational model based on consequentialism would instead have to address the challenge of not advocating actions that could be perceived as horrific, in the name of the greater good (Grau, 2006). This is where Hooker's (2015; 2002, p.86) rule of avoiding disasters could come in play.

The suggested computational model for an IDSS is based on consequentialism in that it calculates the suggestions based on the actions long-term and short-term consequences (where an advantage is given to the long-term consequences). Further, it is inspired by rule-consequentialism in that it has a built-in *avoid disaster protocol*. This comes in play when a certain action's good outweigh the bad, but the potential bad is of the level of a disaster. For example, a truck carrying highly dangerous substances through a populated area. This might be more cost effective and the risk of something bad happening is very slim. Therefore, the good outweigh the bad. But if something were to happen, the consequences would be disastrous, therefore the *avoid disaster protocol* comes in play and the action is not recommended.

The problem of predicting the consequences of an action, especially the long-term, has been the object for criticising consequentialism (Lenman, 2000). This is of course an important notion and therefore the suggested model should incorporate probability in its calculations on suggested actions. But even with this, the suggested actions in the intelligent decision support system should not be viewed as predictions of the future; but as suggested actions based on the likelihood in the available data.

5. Conclusion

The suggestion for computational model for an intelligent decision support system is, to build upon a consequentialist ethical model that, inspired by rule-consequentialism, allows exceptions for extraordinary consequences (disasters), even when the calculated good outweigh the calculated bad. This seems even more important in the crisis management context, where decisions with consequences for human lives must be taken rapidly under stressful conditions. Finally, it must be emphasised that the suggested model is for decision support, and that the main responsibility for decisions still should be carried out by humans, and not by artificial intelligence.

6. Future research

The presented model should have a future implementation that is flexible enough for use in both real crisis situations and crisis management exercises. Before testing the model in sharp real-world crisis management there is a need for a series of iterative evaluations in virtual crisis management exercises. As a complement, these virtual crisis exercises should be enhanced with scripted collaboration. Beside the general collaboration scripts, an interesting specialisation of scripted collaboration are conflict scripts (Mozelius, Borglund & Öberg, 2021). With the future research idea of scripting and evaluating exercise conflicts that resemble moral dilemmas in philosophy.

References

- Alexander, L. & Moore, M. (2020, 30 October). Deontological Ethics. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/ethics-deontological/>
- Annas, J. (2006). Virtue ethics. *The Oxford handbook of ethical theory*, 515-536.
- Brundage, M. (2014). Limitations and risks of machine ethics. *Journal of Experimental & Theoretical Artificial Intelligence*, 26(3), 355-372.
- Cappuccio, M. L., Peeters, A., & McDonald, W. (2020). Sympathy for Dolores: moral consideration for robots based on virtue and recognition. *Philosophy & Technology*, 33(1), 9-31.
- Driver, J. (2014, 22 September). The history of utilitarianism. *Stanford Encyclopedia of Philosophy*. <https://stanford.library.sydney.edu.au/archives/spr2015/entries/utilitarianism-history/>
- Grau, C. (2006). There is no "I" in "robot": Robots and utilitarianism. *IEEE Intelligent Systems*, 21(4), 52-55.
- Gunthner, D. (2012). Emergency and crisis management: Critical incident stress management for first responders and business organisations. *Journal of business continuity & emergency planning*, 5(4), 298-315.
- Hofinger, G., Zinke, R., & Künzer, L. (2011). Psychological requirements for crisis and emergency decision-support systems for public transport control centers. In *ISCRAM*.
- Hooker, B. (2015, 18 November). Rule consequentialism. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/consequentialism-rule/>
- Hooker, B. (2002). *Ideal code, real world: A rule-consequentialist theory of morality*. Oxford University Press.
- Hursthouse, R., & Pettigrove, G. (2016, 8 December). Virtue Ethics. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/ethics-virtue/>
- Kagan, S. (1992). The structure of normative ethics. *Philosophical perspectives*, 6, 223-242.
- Kant, I. (1785). The categorical imperative. In: Warburton, N. (Ed) (2005). *Philosophy: Basic Readings*. Routledge.
- Keenan, J. F. (1995). Proposing cardinal virtues. *Theological Studies*, 56(4), 709-729.
- Koehn, D. (1995). A role for virtue ethics in the analysis of business practice. *Business Ethics Quarterly*, 533-539.
- Konstan, D. (2018, 16 April). Epicurus. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/epicurus/>
- Lenman, J. (2000). Consequentialism and cluelessness. *Philosophy & public affairs*, 29(4), 342-370.
- Misselbrook, D. (2013). Duty, Kant, and deontology. *British Journal of General Practice*, 63(609), 211-211.
- Mozelius, P., Borglund, E. & Öberg, L-M (2021). Scripted Collaboration in Serious Games for Crisis Management Exercises. In *Proceedings of the International Conference on Education and New learning Technologies, Edulearn 2021*.
- Murphy, M. C. (1998). Divine command, divine will, and moral obligation. *Faith and Philosophy*, 15(1), 3-27.
- Powers, T. M. (2006). Prospects for a Kantian machine. *IEEE Intelligent Systems*, 21(4), 46-51.
- Quinn, M. J. (2014). *Ethics for the information age*. Boston, MA: Pearson.
- Ruvinsky, A. I. (2007). Computational ethics. In *Encyclopedia of information ethics and security* (pp. 76-82). IGI Global.
- Segun, S. T. (2021). From machine ethics to computational ethics. *AI & SOCIETY*, 36(1), 263-276.

Simpson, L., (2001). It DOES Matter What You Believe: A Critique of Moral Relativism", CedarEthics Online. 36.

Sinnott-Armstrong, W. (2019, 3 June). Consequentialism. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/consequentialism/>

Velasquez, M., Andre, C., Shanks, S. J., & Meyer, M. (1992). Ethical relativism. *Issues in Ethics*, 5(2).

Weisæth, L., Knudsen Jr, Ø., & Tønnessen, A. (2002). Technological disasters, crisis management and leadership stress. *Journal of Hazardous Materials*, 93(1), 33-45.

Wierenga, E. (1983). A defensible divine command theory. *Nous*, 387-407.

Zinke, R., Hofinger, G., & Strohschneider, S. (2009). Requirements of crisis situations-an action psychology perspective. In *Human factors: a system view of human, technology and organization, Conference Proceedings of the Annual Meeting of the HFES Europe Chapter, Linköping*.