From Games to Moral Agents: Towards a Model for Moral Actions

Joan Casas-Roma, Joan Arnedo-Moreno [1,2]

Abstract. In order to be successfully integrated in our society, artificial moral agents need to know not only how to act in a moral scenario, but also how to identify the scenario first as being morally-relevant. This work looks at certain complex video games as simulations of artificial societies and studies the way in which morally-qualifiable actions are identified and assessed in them. Then, this analysis is used to distill a general formal model for moral actions aimed to be used as a first step towards identifying morally-qualifiable actions in the field of artificial morality. After discussing which elements are represented in this model, and how they are enhanced with respect to those already existing in the analyzed games, this work points out to some caveats that those games fail to address, and which would need to be tackled properly by artificial moral systems.

Keywords. artificial morality, artificial moral agent, moral action, simulation, video games

1. Introduction and Motivations

Artificial morality is a field that aims to furnish artificial systems with moral reasoning capabilities. In the literature, those systems are known as *Artificial Moral Agents* (or AMAs, for short). The reasons why the field of artificial morality is gathering more attention among researchers is directly related to the fact that artificial agents become more and more autonomous, and so more capable of affecting the world with their own decisions, as argued in [15]; autonomous vehicles, military robots and artificial assistants for elderly care are some examples of such systems. Although, as argued in [10], the necessary and sufficient conditions that characterize a *genuine* moral agent are still under debate by philosophers and ethicists, the important thing to note is that, as [1] points out, artificial agents already act in ways that can have moral consequences. Therefore, what is important in this matter is recognizing that autonomous agents that can potentially cause moral good or bad with their actions need to be provided with *some* system that allows them to assess what the appropriate course of action is. Different approaches and challenges have also been discussed in works such as [9], [20], [23], or [16].

Even though they are seldom looked at as such, certain video games embed complex simulations of artificial societies in their virtual worlds. As it has been argued in [8], studying such cases under the lens of multi-agent systems and artificial worlds can provide interesting insights with regards to agent models and simulations; particularly, cer-

[1]The Games Academy, Falmouth University (Cornwall, UK); joan.casasroma@falmouth.ac.uk
[2]Estudis d'Informàtica, Multimedia i Telecomunicació, Universitat Oberta de Catalunya (Barcelona, Spain)

tain games provide detailed representations for the assessment of the moral dimension of in-game actions, which can be studied in conjunction with insights taken from the field of artificial morality. The intersection between morality and video games has received a lot of attention in the literature, although under quite different perspectives: some works such as [19] focus on how to design ethical video games; other works like [24] focus on whether video games can have any impact on the development of their players' moral values; and some works, like [18], or [22], focus on how video games can keep track and evaluate the moral consequences of their characters' actions. Our interest in the present work lies within the last approach —namely, in comparing how certain video games' *morality systems*, which are the mechanisms responsible for identifying and assessing the moral weight of events happening within the virtual world, can be related to insights and challenges that are also present in the field of artificial morality.

This work builds bridges between the fields of artificial morality and the study of complex computer role-playing games, or CRPGs, as artificial societies. The main contribution is to identify and provide a general formal model to represent how morally-qualifiable actions can be identified from an external point of view, inspired by cases taken from different CRPGs. In Section 2, it is argued how, in order to be fully integrated in our society, AMAs would need not only to assess and act accordingly in moral scenarios, but also to be able to identify them beforehand. A study of certain complex video games in Section 3 provides an analysis of how moral actions are identified by the morality systems implemented in those games, which points out to the set of elements that determine when an action has a moral dimension. In Section 4, the insights gained from the previous analyses are distilled into the proposed formal model of morally-qualifiable actions, and some caveats already shown in the formerly reviewed games are discussed. Last, Section 5 presents some conclusions and lines of future work.

2. AMAs Before Moral Action

If artificial morality is to be effectively integrated at a large scale in AI systems, those would need to be equipped not only with the capacity of acting in a morally-desirable way when faced with a clearly-identified moral scenario, but also to be able to identify morally-relevant scenarios by themselves, as well as assessing the relevant factors that evaluate those scenarios as being either good or bad. Current approaches to the design and implementation of AMAs, such as [2], or [16], rely on having relevant moral situations already identified and on spoon-feeding them to the systems. Then, the AMAs need to assess what the moral value of that scenario is, and decide how to react accordingly. If moral reasoning is to be integrated in the overall cycle of artificial systems, though, one needs to consider a bigger picture when it comes to recognizing the phases of moral action. In particular, an AMA needs to:

1. Identify: Prior to bringing moral reasoning into the decision system, an AMA needs to be able to distinguish between events that have moral relevance, or not.
2. Assess: Once the AMA has identified a certain event as having a moral dimension, it must be able to assess whether that event can be considered morally good, or bad, and compute "how much" good or bad it is.

3. React: Once the agent "understands" the moral value of the event it is evaluating, it can decide how to react accordingly and choose an action aiming towards the most desirable outcome[3].

Current existing approaches to AMAs focus on the last two steps: namely, assessing moral events and reacting to them. However, in those cases the relevant events are already given to the AMA as pre-identified morally-qualifiable events. Due to this, this work draws the attention towards the first step of identifying moral events, which would need to be integrated in any existing AMA that is meant to incorporate moral behavior not only as an output for pre-identified events fed into it, but as one of its built-in features in relation to its environment. In order to understand how morally-qualifiable events can be identified from the perspective of an external observer, this work looks through a somewhat unusual lens: the study of CRPGs as simulations of artificial societies.

3. Video Games as Simulations of Artificial Societies

Some CRPGs present the player with a complex virtual world inhabited by virtual agents (often referred to as non-player characters, or NPCs) which often follow their own schedules, routines, needs and desires across the in-game world, and do so independently from the actions of the player. Furthermore, some of these CRPGs capture a high degree of detail with respect to the way those NPCs react to the actions carried out by the player, depending on whether these actions are deemed to be "good" or "bad" with respect to the game's morality system. In order to do so, those games usually feature a comprehensive system to build and keep track of the player's moral persona within the game as a reflex of the player's behavior towards the inhabitants of the virtual world. As it is argued in [8], some of these games present an architecture for morality systems in virtual societies that could lead to promising research lines, if combined with the fields of multi-agent systems and artificial societies.

Aside from pre-scripted choices that are part of the main plot, most open-world CRPGs also offer a high degree of freedom of action allowing for the appearance of "off-script" emergent play, which is a highly sought out property in open-world games. It is precisely with regards to this kind of actions that the game needs to implement some sort of morality system to recognize which of those are morally relevant, assess whether they are considered good or bad by the in-game value system, and reflect that on the player's moral persona. When considering how this kind of games identify and assess morally-qualifiable actions, the game engine can be seen as a sort of "overseer" with nearly perfect information about the state of the game's world and its virtual agents; by studying this kind of games through this lens, this work revolves around the following questions:

1. How does the game engine identify morally-relevant actions?
2. What are the relevant elements used to assess their moral weight?
3. Can different ethical theories be captured by those morality systems?
4. Is there anything relevant that cannot yet be captured in such setting?

[3] Works like [12] argue that a decision on a moral action should never be left to the machine alone –that is, there should always be a human-in-the-loop. This topic, however, is beyond the scope of this paper.

Answering the first and the second questions leads to the definition of a general model for morally-qualifiable actions; exploring the third question shows how existing morality systems already allow to capture the particularities of different ethical theories in a computational setting; finally, the fourth question points out to further challenges for the field of artificial morality, as it is argued how certain features of moral actions cannot yet be captured, even in a setting with an overseer with almost perfect information. The following subsections provide an overview of three games with morality systems that capture three different ethical theories, and which are used to discuss the previous questions.

3.1. The Elder Scrolls and the Consequentialist Approach

A game could be designed to represent some form of Jeremy Bentham's account of *utilitarianism* [14], which aims to measure the amount of happiness and pain resulting from an action, and which defends the ethical principle that the "best" action is the one that maximizes happiness for the maximum amount of people. Even though this approach is not devoid of complications, it provides the designers with a straightforward way of expressing moral weight in mathematical terms. As the game engine has access to the data of everything belonging to the virtual world and its inhabitants, it is relatively easy to measure how much certain events affect this data, and to interpret those changes as being good or bad for the world and the artificial characters living in it.

Although games do not usually implement a straightforward calculation of happiness and suffering in their morality systems, some form of consequentialism is, probably, the easier way to account for the morality of the player's actions. An example of a well-known series of games conforming to this is *The Elder Scrolls* saga, in which a particularly detailed morality system can be found in the fourth title of the saga, *The Elder Scrolls IV: Oblivion* [3]; this system does not only account for the moral persona of the player, but also make NPCs more or less prone to respect the in-game law system, depending on their urge to satisfy certain needs (see [7] for more details on this, as well as a proposal on how to enhance the existing in-game mechanics to furnish non-playable characters with a certain degree of operational moral autonomy).

What makes an action good or bad, in this kind of games, is usually not the action in itself; instead, the relevance falls on the "patient" that receives that action. For instance, killing a character identified by the game as non-evil, such as a peaceful citizen, normally results in the player character losing "moral points", whereas killing an evil NPC, such as a bandit, grants positive moral points. Considering this, it can be seen how, by moving the point of evaluation of a moral event to the patient, rather than the action itself, games following this approach fall into a consequentialist approach that aims to evaluate the balance between the good and bad outcomes of an event, rather than aiming at an act-centered deontological approach that would classify events depending solely on the act itself.

3.2. Fallout and the Intentionalist Approach

Other ethical systems, inspired by Kantian ethics, could also be considered. In a nutshell, those accounts determine what is moral and what is not by looking at the "will", or "intention" behind an action. What matters, rather than the actual outcome of a choice, is whether the intention of the one who carried out the action was good. However, and as

pointed out in [11], it would be very difficult to *effectively* implement this kind of ethical system within a game, as the game would need to reach beyond the player character, and straight into the player's intentions. However, the games in the *Fallout* series base the main measurement of the player's moral profile on the notion of *karma*, which aims to account not for the consequences of the player's choices, but rather for the intention behind those choices (see [6]). But how do those games aim to account for that, given what it has just been said about capturing intentions?

In order to illustrate that, let us consider a case from *Fallout 3* [4], which is no strange in the academic literature: the *Tenpenny Tower* quest line. Briefly explained, this quest takes the player to a tower inhabited by a group of humans, and where a band of ghouls (which are horribly-mutated humans as a result of exposure to radiation) also want to live in; nevertheless, the tower residents do not want the ghouls to live in there. The player is given three ways of resolving this quest: 1) to side with the humans and kill the band of ghouls; 2) to side with the ghouls and help them sneak into the tower (which would result in them killing the human inhabitants); 3) or to find a diplomatic solution to the conflict, allowing both humans and ghouls to live together in the tower. The "good" choice, karma-wise, is to find a diplomatic solution to the conflict. After doing so, the player can come back to the tower a few days later to find both humans and ghouls apparently getting along pretty well; nevertheless, coming back after a few more days shows that the humans and the ghouls had a "disagreement" and the ghoul leader decided to "take out the trash" —at the tower's basement, the player will then find the bodies of all the former human inhabitants.

However, and even though the long-term consequences of the diplomatic solution to this quest can be considered bad, the game rewards the player with good karma when following this approach. The reasons behind this is, precisely, because the intention behind the player's choice was good, the consequences that would likely follow from such choice were meant to be good, and the resulting slaughter was a long-term, unforeseeable consequence that goes beyond the player's actual involvement in the conflict. This depth of unforeseeable, player-independent consequences on the player's choices is, precisely, what makes moral choices in Fallout so unique with respect to other video games. By reflecting on the way Fallout aims to capture intention in the player's actions, it can be seen how this notion, when measured from the point of view of an external system, can be defined as:

> The *intention* of an action can be externally identified as the consequences that would most likely follow from that action, according to the state of affairs believed by the agent at the moment of performing the action.

This definition has strong parallelisms with the well-known *Belief - Desire - Intention* (BDI) theory (firstly defined in [5]); although going deeper into this parallelism would fall outside the scope of the present work, this "vanilla" version of intention that has just been defined could be used to account for such notion from an external point of view, thus tackling the other-minds problem, and being a starting point to represent it in artificial systems[4]. Regardless of this particular example to show how Fallout 3 aims to capture the player's intentions, the game follows the same reasoning regarding other

[4]It is only fair to note that, by tapping unto the agent's beliefs in order to account for intentions, the other-minds problem is, in the end, only moved from one place to another.

actions, meaning that the moral alignment of the patient mainly determines whether the action is deemed to be good or bad by its morality system.

3.3. Ultima and Virtue Ethics

Conversely, a game could adopt *virtue ethics*, founded by Plato and Aristotle and defended by contemporary authors such as [13], as its underlying ethical system. Probably the most well-known saga that takes virtue ethics as their underlying ethical system is the *Ultima* saga, with a special emphasis on *Ultima IV* [17]. In contrast with utilitarianism, or intentionalism, virtue ethics is based on the development and cultivation of habits and behaviors in accordance to a certain set of virtues; given the medieval fantastic setting of the game, which is inspired by medieval chivalry codes, those virtues include compassion, justice, or valor, to name a few. In this case, a morality system capturing virtue ethics as the underlying ethical system should focus on which particular virtue an action would account for, instead of just measuring the consequences of such action, or the intention behind it.

In Ultima IV, virtues are mapped into in-game actions by identifying how those actions align to one or another virtue, depending on a set of factors. For the sake of simplicity, the following example focuses only on how two different virtues[5], *compassion* and *valor*, work with respect to a single action, fighting. Furthermore, the example focuses on two possible outcomes in a fight: to flee from it, or to kill the enemy. What is interesting in this case is that the player fleeing from the fight, letting the enemy flee, or killing the enemy have quite different effects on the virtues of compassion and valor, depending exclusively on one feature: the enemy's moral alignment. NPCs are tagged within the game as being either *evil*, or *non-evil*. Now, if the player is engaged in a fight with an evil NPC, killing it will increase the virtue of valor for the player, whereas fleeing will decrease the player's valor. Conversely, if the player is engaged in a fight with a non-evil NPC, such as a peaceful citizen, or even a wild animal such as a deer, just the act of engaging in a fight with it will result in a severe penalty to the virtue of compassion, whereas fleeing from the fight, or allowing the NPC to flee, will result in a compassion benefit. Table 1 summarizes this brief example.

	Evil NPC	Non-evil NPC
Fight	↑ Valor	↓ Compassion
Flee	↓ Valor	↑ Compassion
Allow to flee	-	↑ Compassion

Table 1. Actions, patients and virtues in Ultima IV.

Just as it happened with the previous games, the only thing that determines the virtuous outcomes of how the fight resolves has to do with the patient that "receives" those actions and, in particular, with the patient's alignment. In this case, though, instead of just polarizing the actions towards one side of a good / evil axis, Ultima IV goes one step beyond and defines a complex net of relations between agent, action, patient and virtues, thus showing how existing morality systems in video games can also capture virtue ethics defined in a pretty detailed way by means of a top-down system.

[5] Virtues in Ultima IV is quite complex, but only those two virtues are considered here for simplicity. For more information, see https://strategywiki.org/wiki/Ultima_IV:_Quest_of_the_Avatar/The_Shrines_of_Virtue .

4. Towards a General Model for Actions

As is has just been shown, several CRPGs featuring an open-world with a certain freedom of action need to include a way of identifying whether those actions are considered good, or bad, within the game. This is done both to account for short-term reactions, such as how the NPCs affected by those actions behave, as well as to account for a longer-term characterization of the player's moral persona. Even though different games account for different nuances, it can be seen how there are three major factors that are taken into account in all of them: the *agent* that carries out the action, the *act* characterizing it, and the *patient* who receives its consequences.

Taking all this into account, the schema depicted in Figure 1 represents a general model for the identification of morally-qualifiable actions. It should be noted that this model, as it is, allows to identify events that have a moral weight; however, in order to assess the weight of those events, one needs to have an underlying set of moral values that determine what is considered to be good, or bad, such as the previously studied games did. As expected, different sets of moral values, or different priorities among those, will lead to the same event as being morally-qualified in a different manner.

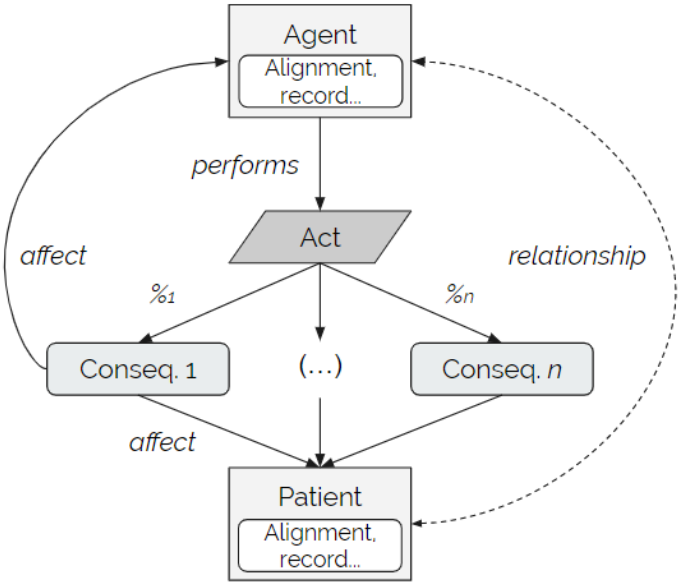

Figure 1. Relevant elements in the identification of a moral action.

The structure of moral actions depicted in Figure 1 captures the most relevant elements identified by the cases analyzed in Section 3, and accounts for:

- An *agent*, for whom we have information about its *alignment*. Potentially, instead of a label specifying the agent's alignment, the model could keep track of the agent's *record* of morally-qualifiable acts which, in turn, could intrinsically represent the alignment.
- The *act* performed by the agent, which has a set of *consequences* $[1 \dots n]$ following from the act and given a certain probability, and which affect the patient of the

moral action. Following the insights discussed in Section 3.2, one can externally interpret the agent's intentions, with respect to an act, in terms of the most likely outcomes that such act could have.
- A *patient* who receives the consequences of the act[6], and who also has an associated moral alignment, or a record of morally-qualifiable actions intrinsically representing the patient's values.

In addition to the properties identified in the analyzed cases, the model has been enhanced to capture two additional properties that can be relevant not for the identification of a morally-qualifiable action, but rather for its assessment:

- The *relationship* that the patient and the agent may have, and which could be relevant in terms of understanding why the agent has decided to perform the act affecting the patient. Such feature could capture, for instance, the fact that assuming that one of the patients in the trolley dilemma is a relative of the agent in charge of deciding whether to pull the lever can influence the decision made by the agent.
- The way consequences of the act can also *affect* the agent, aside from the patient. This could allow the model to take into account both a theory of right and a theory of good (see [21]), which would allow to distinguish situations featuring an act that could be seen as negative from the point of view of a theory of right, such as stealing food in a market, but which could be seen as acceptable, from a theory of good, if the agent who steals the food is a poor, starving child.

In order for this model to be implemented into an AMA, some details would still have to be refined. For instance, determining what entities qualify as patients (human beings, living organisms, the environment, artifacts, etc.), or whether the same consequence has the same effect on each patient, would be central issues that would need to be addressed. In addition, the AMA would need to play a double role in this identification, as it would be both the overseer of the action, as well as potentially the source of it (i.e.: the agent). This is, however, an endeavor we leave for future work: as we point out in the next section, there are certain features that neither the games we took as inspiration, nor the model we provide, can account for yet, and which would need to be addressed before thinking about implementation details.

4.1. What is Missing?

Despite the fact that some morality systems can show a good degree of granularity, there are still some important details that have not yet been captured in games, and which can play a crucial role when assessing the moral weight of an event. It could be argued how if those caveats arise from such an ideal setting, they will probably do as well when trying to implement a way of identifying and assessing moral events in the real world, where information is deemed to be incomplete, inaccurate, and sometimes even incorrect.

The first shortcoming is the inability to link different events: an example from Oblivion can be used to show where the problem lies. In general, if one considers the in-game action of killing a character, the game reacts as expected, with regards to the alignment of the patient: if the player kills an evil NPC, she is rewarded, but if she kills a non-evil

[6]Some consequences could likely affect more than one patient; therefore, and although only one is represented in Figure 1 for the sake of simplicity, multiple patients should be taken into account.

NPC, she is punished... unless the non-evil NPC attacks first, in which case the player is "justified" and can kill it without any kind of moral penalty. This effect leads to a nasty strategy consisting in casting a spell that puts the NPC into an "enraged mode" that makes it attack everyone in sight, which then gives "free way" for the player to kill it without any kind of moral consequence. In this case, it can be clearly seen how the game lacks the ability to identify how casting this spell towards a non-evil NPC is clearly something that could count as a morally-qualifiable action, specially when it leads to the player killing the NPC in the end. As such, artificial morality systems should be able to trace a cause - effect chain between seemingly independent actions and see how the state of affairs brought about by one had affected the conditions for the following.

From the inability to recognize the former issue, one can get a free pass on Machiavelian scheming to achieve one's ends through planning morally-shady independent events that lead to a desired outcome. As an example, it can be shown how, although attempting to capture the players' intentions behind their actions, Fallout 3 still fails to recognize such cases. Still in the same quest mentioned in Section 3.2, the game allows the player to follow certain shady ways to get, for instance, the ghouls killed without any karma loss by not getting directly involved in it. This can be done by luring hostile NPCs into the ghouls' den, which results in the ghouls getting killed, but not as a result of the player's direct action. Considering this, it is interesting to note how, even if the notion of karma in Fallout does indeed take intention into account up to a certain degree, it ends up still being dependant on mostly isolated choices and actions that are directly performed by the player, but which may not account for a more far-reaching scheming that pursue darker ends, based on more nuanced, long-term intentions, and which involve chaining different, seemingly independent events.

By highlighting this two interrelated caveats, it can already be seen how a proper artificial morality system would need not only to be able to identify and assess morally-qualifiable actions on their own, but also to keep track of possible causality chains between seemingly independent actions, as well as having a way of determining how the state of affairs brought about by one action could have played a role in the assessment of the moral weight of a later action. This, therefore, should be a priority for researchers working on topics such as the design and implementation of artificial moral agents, or the definition of formal models of moral actions.

5. Conclusions and Future Work

By looking at complex role-playing games as simulations of artificial societies, this work identifies a set of features used by the morality systems implemented in those games to identify and assess morally-qualifiable actions. Those insights are then abstracted from the context of the video games and distilled into a proposed formal model highlighting the features needed to identify moral events from an external perspective. In doing so, this work builds bridges between the fields of artificial morality and video games understood as simulations, and it tackles the question of what comes *before* artificial moral agents need to act; that is, how to distinguish morally-qualifiable actions from actions with no clear moral content, and how to assess their moral weight. In this sense, the proposed model is a valuable contribution not only to the field of artificial morality, but also potentially as a way of further refining the simulation aspect of complex video games.

When discussing the proposed model, certain shortcomings that can already be found in video games' morality systems are identified; those caveats concern the lack of a joint, holistic understanding of separate events that, when put together, may reveal an underlying goal that was not shown explicitly by any of the individual events; this point would need to be tackled by considering chains of cause-effect by keeping track of previous actions, and how certain consequences of an action *a* bring about conditions for action *b* to happen. Once these caveats are solved, the model should be adapted in order to allow implementation in an AMA: linking the models' consequences and patients into the AMA's representation of its environment pose a challenge that would need to be addressed as well.

References

[1] I. Allen, C.; Smit and W. Wallach. Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and Information Technology*, 7(3):149–155, 2005.

[2] M. Anderson, S. L. Anderson, and C. Armen. An approach to computing ethics. *IEEE Intelligent Systems*, 21:56–63, 2006.

[3] Bethesda Game Studios. *The Elder Scrolls IV: Oblivion*, 2006. Game (PC, Playstation 3, Xbox360).

[4] Bethesda Game Studios. *Fallout 3*, 2008. Video game (PC, Playstation 3, Xbox 360).

[5] M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.

[6] J. Casas-Roma and J. Arnedo-Moreno. Characterizing morality systems through the lens of fallout. In *Proceedings of the Digital Games Research Association Conference*, DiGRA '19, 2019.

[7] J. Casas-Roma, M. J. Nelson, J. Arnedo-Moreno, S. E. Gaudl, and R. Saunders. Towards enhancing npcs' morality: The case of the elder scrolls iv: Oblivion. In *Artificial Intelligence and Simulated Behaviour Conference*, AISB '19, 2019. 16th – 18th April (Falmouth, UK).

[8] J. Casas-Roma, M. J. Nelson, J. Arnedo-Moreno, S. E. Gaudl, and R. Saunders. Towards simulated morality systems: Role-playing games as artificial societies. In *Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, volume 1 of *AIIDE '12*, pages 244–251, 2019.

[9] L. Floridi and J. W. Sanders. On the morality of artificial agents. *Minds and Machines*, 14(3):349–379, 2004.

[10] D. J. Gunkel. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press, 2012.

[11] M. Heron and P. Belford. 'it's only a game' — ethics, empathy and identification in game morality systems. *The Computer Games Journal*, 3(1):34–53, 2014.

[12] R. López de Mántaras and P. Meseguer. *Inteligencia artificial*. CSIC, 2017.

[13] A. MacIntyre. *After Virtue: A Study in Moral Theory*. University of Notre Dame Press, 2007.

[14] J. S. Mill. *Utilitarianism and other essays*. New York: Penguin Classics, 1987.

[15] C. Misselhorn. Artificial morality. concepts, issues and challenges. *Society*, 55(2):161–169, 2018.

[16] I. Muntean and D. Howard. Artificial moral agents: Creative, autonomous, social. An approach based on evolutionary computation. In J. Seibt, R. Hakli, and M. Norskov, editors, *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy 2014*, pages 217–230. IOS Press, 2014.

[17] Origin Systems. *Ultima IV: Quest of the Avatar*, 1985, 1990. Video game (multiple systems).

[18] A. Russell. Opinion systems. In *AI Game Programming Wisdom 3*. Charles River Media, 2006.

[19] M. Sicart. Moral dilemmas in computer games. *Design Issues*, 29(3):28–37, 2013.

[20] J. P. Sullins. Ethics and artificial life: From modeling to moral agents. *Ethics and Information Technology*, 7:139, 2005.

[21] M. Timmons. *Moral theory: an introduction*. Rowman & Littlefield Publishers, 2012.

[22] J. Švelch. The good, the bad, and the player: The challenges to moral engagement in single-player avatar-based video games. In K. Schrier and D. Gibson, editors, *Ethics and Game Design: Teaching Values through Play*, pages 52–68. IGI Global, 2010.

[23] F. S. Wallach, W. and C. Allen. A conceptual and computational model of moral decision making in human and artificial agents. *Topics in Cognitive Science*, 2(3):454–485, 2010.

[24] J. Zagal. Ethically notable videogames: Moral dilemmas and gameplay. In *Proceedings of the 2009 DiGRA Conference*, 2009.