ICRES 2019: International Conference on Robot
Ethics and Standards, London, UK, 29-30 July 2019.
https://doi.org/10.13180/icres.2019.29-30.07.id#

COMPETING ETHICAL FRAMEWORKS MEDIATED BY MORAL EMOTIONS IN HRI: MOTIVATIONS, BACKGROUND, AND APPROACH

RONALD C. ARKIN, JASON BORENSTEIN
Georgia Institute of Technology, Atlanta, GA, USA 30332-0280
arkin@cc.gatech.edu, borenstein@gatech.edu

ALAN R. WAGNER
Robot Ethics and Aerial Vehicles Lab, Penn State University
University Park, PA 16802-7000 USA
alan.r.wagner@psu.edu

This paper describes ongoing research for a three-year NSF-funded project on ethical architectures for robots; specifically striving to understand how robots can reconcile differing outcomes produced by alternative ethical frameworks (e.g., Kantianism, Utilitarianism, and Ross's moral duties). The process of determining the correct action is mediated by context and the moral emotional state of the robot. This paper describes the motivation, background, and approach to the project.

1. Introduction

Ethical decision-making is fraught with difficulty, certainly for robots let alone humans. Human reasoning is not static; for instance, an individual may decide to act differently when faced with the same situation multiple times. If a robot's ethical decision-making process is going to be designed based on some approximation of how humans operate, then the assumption is that a good model of how humans make decisions is readily available. Yet many complexities need to be addressed regarding the nature of ethical decision-making including:

- Is there a singular ethical framework that does or should guide decision-making?
- Which is a more important aim: respecting rights or the pursuit of positive outcomes? And does the relevant answer depend on the circumstances?
- Should the main benefits of a robot's actions accrue to society or the individual with whom the robot interacts?
- How does cultural or social context play a role in ethical decision-making?
- Are there occasions where it is more important to adhere to social norms instead of doing what in principle may be the legally correct thing to do?
- What role do moral emotions (e.g., shame, guilt, and empathy) play in these decisions?

This paper describes recently initiated and ongoing research attempting to answer these questions. Towards that end we are investigating how to create a robot that is equipped with multiple ethical reasoning systems so that it can do the right thing. But what is doing the right thing? For example, should ethical rules be bent and if so, under which circumstances? Is deception acceptable if it results in better outcomes for most or all concerned? We explore the reasoning process that a robot should use when it is tasked with making an ethical choice. We intend to employ two strategies: (1) to evaluate whether a robot acts in a way that ethical experts would endorse; and (2) whether the robot mimics average human behavior in similar circumstances.

The ethical frameworks that could, in principle, be encoded in a robot include deontological (Kantian/Rights-based) methods, consequentialist (utilitarian) approaches and social justice (Rawlsian principles, Ross's moral duties) frameworks. Yet these frameworks can lead to

conflicting recommendations. In our approach, an ethical decision will be mediated by the emotional, or simulated emotional, state of the agent (human or robot) and we use these moral emotions to select which framework should take dominance for a particular situation. These are to be compared to the action selections of humans in similar situations.

This research leverages our considerable experience in ethical decision-making in health care and military scenarios. Here, we focus on a new domain (game playing) and build on our work in preserving dignity in situations where power dyads exist (e.g., teacher-student, parent-teenager, and caregiver-patient). As such, the underlying computational architectures we have developed are expanded to reason in multiple ways according to different ethical frameworks; the selection of the framework and resulting action will depend upon the moral emotional state of the agent.

Experimental humanoid testbeds are being developed at Penn State using the Pepper robot, and at Georgia Tech using the smaller Nao and Milo robots. The research domains that have inherent divergent ethical choices include:

1. Game playing with a child and the role of other-deception for losing on purpose based on the perceived emotional state of the child.
2. In eldercare, looking at tasks such as pill sorting using deception as a means to reduce frustration to facilitate training with the associated trade-off on safety.

Experimental methods will involve recording human decision-making as a baseline, evaluating each architectural framework independently, and then comparing to a flexible ethical action integrated architecture with multiple frameworks mediated by the moral emotional state of participants. The key goal of the robot's computational architecture is to reproduce a typical human ethical choice (which for simplicity's sake will be referred to as "folk morality") and/or a choice that reflects the consensus of ethical experts where available. Specifically, the following research questions are being addressed:

1. Whether a computational architecture that analyzes a situation from multiple ethical frameworks can enable a robot to do the right thing in a particular context.
2. Whether a robot's ethical performance improves with experience with respect to ethical expert consensus or folk morality (normative ethical reasoning).
3. Whether the approach is dynamic and adaptable to reflect different environments and situations.
4. Whether the robot makes better ethical decisions when guided by multiple ethical frameworks instead of just one.

2. Related Work

The field of related work on approaches to creating an ethical autonomous system is extensive [e.g., 1-4]. Yet three primary methods have been proposed. One method is to have an autonomous system model the behavior of an ethically competent exemplar [3]. Inverse reinforcement learning might serve as means for framing such learning [5]. While the possibility of using inverse reinforcement learning, or some other means, to model the behavior of an ethical exemplar has been considered, this kind of approach raises a number of important concerns such as the introduction of cultural biases and the potential lack of adaptability. While the autonomous system could use an ethical exemplar to learn some subset of appropriate behavior, it is not clear how the agent or robot would adapt what it has learned to novel situations and contexts.

Some scholars suggest preprogramming legal and ethical rules into such a system [6], and by following such rules, an autonomous system might perform ethical actions within some well constrained environments (e.g., [7,8]). This has the clear advantage that these preprogrammed rules are agreed upon to be philosophically and legally grounded. Moreover, these rules have some level of explainability in that the autonomous system can simply point human operators

or interactive partners as the basis for the rule's history or origin, in a military context, this could, for example, be the Geneva Conventions.

Others have explored the possibility of using an ethical theory as an underpinning for an autonomous system's ethical reasoning [9]. Some philosophical theories of ethics (e.g., Utilitarianism) more easily lend themselves to software encoding and robot action selection than others. While many researchers have investigated both formal and ad hoc methods for encoding ethical frameworks for use by an autonomous system, our approach generates action recommendations by drawing on several ethical frameworks [10-12]. The autonomous system then chooses the action that best fits the situation. This flexibility may allow the system to be more adaptive when confronting a situation that it has not faced in the past. The section below briefly describes several philosophical traditions in the realm of ethics that inform the design of the project's computational architecture.

3. Ethical Frameworks

A wide range of approaches in the realm of ethics seek to provide insight in terms of what counts as an ethically appropriate or inappropriate act. Some of these approaches rise to the level of an ethical theory in the sense that philosophers have endeavored to provide a foundation that guides all ethical decision-making (e.g., Kantianism or Utilitarianism); whereas other scholars have articulated a scaffold that gives general guidance for at least some types of decisions or policies without necessarily providing enough nuance and specificity to address all types of ethical concerns (e.g., the Capabilities Approach [13]). For our purposes, we use the term "ethical framework" to broadly encompass both types of approaches in ethics. Many ethical frameworks could potentially be used as a guide for decision-making. For this project, we consider encoding at least three ethical frameworks into a robot's design matrix as a basis for action: (1) Utilitarianism, (2) Kant's ethical theory, and (3) W.D. Ross's duty-based view.

Consequentialism refers to ethical approaches that seek to define ethical goodness and badness in terms of the consequences that an act produces. Utilitarianism is a widely-embraced type of consequentialism; it is a theory that often influences the framing of public policy decisions. Many versions of Utilitarianism have emerged, including Act, Rule, Ideal, and Preference Utilitarianism [14]. While scholars disagree on key details of each version, what unifies them is the importance of the theory's fundamental maxim to pursue "the greatest good" for society. One method is to undertake stakeholder analysis whereby the potential benefits and harms that a course of action may have on different entities are assessed.

Deontology is a collection of views that converge around the notion that ethical rightness and wrongness is determined independently of the consequences of one's actions and instead is defined by whether ethical obligations have been upheld [15]. A key tenet of deontological reasoning is that moral imperatives serve as rules governing decision-making across many or most situations. The cornerstone of Immanuel Kant's view, for example, is his Categorical Imperative, a formula that is supposed to guide decision-making towards absolute and universal ethical actions.

There are many other rights and duty-based approaches within the realm of deontology, including ones that reject the absolutism contained within Kant's view. For example, Ross sought to glean insights from Utilitarianism and Kantianism, and incorporate them into view that captures the ethical obligations that humans have to one another [16]. He articulated a collection of *prima facie* duties; each ethical duty is something we must uphold but one duty could be superseded by a more important duty in a particular case.

3.1 Factors Contributing to Ethical Disagreements

Countless factors can influence how ethical decisions are made [17] and such factors can contribute to the emergence of ethical disagreements; for example, varying emotional responses to the same situation, one person responds with anger as compared to another who experiences joy, can lead to a dispute. Implicit bias and other psychological aspects can also contribute. In addition, logical fallacies such as an appeal to (an unqualified) authority can play a key role. In the next sections, we focus on how the reasoning process can lead to disagreements on ethics.

3.2. Disagreements within an Ethical Framework

In principle, moral reasoning should lead to the same conclusion about what counts as an ethical action. Yet if one were to assume that a singular ethical framework correctly characterizes how ethical decision-making should work, then a question emerges regarding how advocates within the framework account for disagreement. Kantians argue that certain ethical obligations, such as the duty to avoid lying, are absolute and that if the Categorical Imperative is correctly applied, then all rational agents should in principle arrive at the same answer to a moral problem. Thus, it is difficult for Kantians to account for the problem of moral conflict [18]. However, a Kantian could suggest that ignorance or a misapplication of reason could generate disagreement about the ethical course of action. Utilitarians converge on the notion that the principle of utility must be pursued but conflict can occur between those applying Utilitarianism. What can contribute to conflicting Utilitarian views is weighing a certain consideration (e.g., the number of stakeholders affected) more heavily than another (e.g., the duration of the effect). Alternatively, Utilitarians might disagree on how narrowly or broadly the scope of ethical analysis should be framed for a particular ethical issue.

3.3 Disagreements Across Ethical Frameworks

Ethical disagreement can also emerge because a lack of consensus persists about which particular ethical framework should be embraced and which values and goals need to be prioritized. For example, arguments continue between Kantians and Utilitarians about what the fundamental goal of ethics is. The former emphasizes that respect for persons as the primary aim whereas the latter focuses on producing goodness for society.

3.4 Resolving *an Ethical Disagreement*

There are many ways in which a resolution to an ethical disagreement can be reached; we cannot fully address that here. We provide a few brief examples for illustrative purposes. Ethical disputes are sometimes resolved in non-rational or irrational ways such as through intimidation or coercion. Yet the overarching hope is that "good reasons" will prevail and are used as a basis for identifying the ethically appropriate course of action [19]. For instance, whether "good reasons" can be found for prioritizing one's own self-interest over the interests of others might help resolve a moral conflict. Finding a rational solution is often tied to whether important values, such as fairness and honesty, have been upheld.

For the purposes of our research, moral emotions are used as a foundation for guiding decision-making when ethical disagreement emerges. More specifically, the moral emotions of those people involved can help enable a robot to navigate situations where disagreement emerges between/among two or more ethical frameworks about the appropriate course of action. We leave aside circumstances where the debate resides within one framework (e.g., between Utilitarians) but the computational architecture developed could potentially be used to handle such situations as well.

4. Moral Emotions

The relatively young machine ethics community has, to date, largely focused on developmental ethics, i.e., how an agent would develop its own sense of right and wrong in situ. In general, many of these efforts typically ignore moral emotions as a scientific matter worthy of consideration. Nonetheless, considerable research has been conducted on the role of emotions in robotics, including work in Arkin's laboratory over the past 20 years [20]. Far less explored in robotics is the set of moral secondary emotions, and their role in robot behavior and human-robot interaction. De Melo et al. [21] demonstrated that presence of moral affect in human-robot interaction is both discernible and enhances interplay between humans and robot-like avatars.

Our research [22] in the moral affective space research is illustrated by the use of guilt incorporated into an ethical robotic software architecture. Guilt is "caused by the violation of moral rules and imperatives, particularly if those violations caused harm or suffering to others" [23] and is capable of producing proactive, constructive change [24]. The specific architectural component we have implemented, referred to as the ethical adaptor, incorporates Smits and De

Boeck's [25] mathematical model of guilt; it is used to proactively alter the robotic system's behavior in a manner that can lead to a reduction in the recurrence of a guilt-inducing event. Simulation results demonstrate the ethical adaptor in operation [22].

There is ample evidence that moral emotions guide human beings in making ethical judgments. As per Cameron et al. [26] "Feelings were long dismissed as unworthy of study, especially in morality and ethics … but decades of research … reveals that they matter for moral judgment. Emotional feelings … can both intensify and diminish moral judgments, but the precise link between feelings and moral judgment is debated." (See [26] for 18 supporting references denoted by ellipses deleted due to space limitations.) Using previous methods [22], moral emotions can bias the system behavior in a way that supports positive moral judgment and resulting behavior, (e.g., reducing the likelihood of re-occurrence of an immoral act) and to maintain a partial theory of mind representation of the affective state of the human counterparts, acting to foster their emotional state consistent with enhancing the dignity of the people involved. We consider situations involving the care of older adults or children, with the goal of preserving human dignity in that relationship.

Arguably, in order for an autonomous agent to be truly ethical, emotions may be required at some level: "*While the Stoic view of ethics sees emotions as irrelevant and dangerous to making ethically correct decisions, the more recent literature on emotional intelligence suggests that emotional input is essential to rational behavior"* [27]. Gazzaniga [28] identifies three neuroscientific aspects of moral cognition: (1) moral emotions, which are centered in the brainstem and limbic system; (2) theory of mind, which enables us to judge how others both act and interpret our actions to guide our own social behavior, where mirror neurons, the medial structure of the amygdala, and the superior temporal sulcus are all implicated in this activity; and (3) abstract moral reasoning, which uses many different components of the brain. Moral behavior, we firmly believe, involves the use of moral emotions, guiding intuitions in determining ethical judgments, although this is not universally agreed upon. Haidt [23] provides a taxonomy of moral emotions: Other-Condemning (Contempt, Anger, Disgust); Self-Conscious (Shame, Embarrassment, Guilt); Other-Suffering (Compassion); Other-Praising (Gratitude, Elevation). We allow these emotions to bias the behavior of the system (e.g., as in [22]), and as appropriate, maintain a partial theory of mind representation of the affective state of the robot's human counterparts in order for the robot to act in a manner enhancing their dignity.

5. Architectural Overview

There is an important role for artificial emotions in personal robotics as part of meaningful human-robot interaction. It is clear that value exists for their use in establishing long-term human-robot relationships and in our case for supporting ethical decision-making. Our research addresses moral affective aspects of the system with respect to mediating ethical behavior. This secondary class of emotions has been mostly ignored by many within the robotics and computer science community, with a few exceptions (e.g., [21]). For example, once it has been determined that the robot's actions involve an empathy-inducing situation, it is necessary to compute the appropriate magnitude of empathy that should be expressed for a specific situation in order to assess the level of bias towards one ethical action or another. We use the same methods employed for our models for negative moral emotions (i.e., guilt [22]) to look at positive moral emotions and abstract these into models suitable for computational implementation. The ethical adaptor uses a modified version of the Smits and De Boeck model [25] to compute the system level of a moral emotion. Using this first approximation, we express a positive moral emotion in terms of situational appraisal values, norm appraisals, other-evaluations, evaluations about the act that elicited the emotion, and motivation and action tendencies geared towards other-support. The model then assigns the probability for "feeling an emotion" as: where $P_{i,j}$ is the probability of person i feeling the emotion in situation j, $\text{logit}(x) = \ln(x/(1-x))$, $\beta_{j,k}$ is the emotion-inducing power of component k in situation j, θ_i is the emotion's threshold of person i, σ_k is the weight of component k contributing to the emotion, τ is an additive scaling factor,

and a_j is a scaling weight for situation j. In particular, again using empathy as an example, instead of computing the probability that empathy results from some situation, the ethical adaptor computes the magnitude of empathy that robot i should experience in situation j as: $Empathy\ (i,j) = a_j(\beta_j - \theta_i)$. In the current implementation of the ethical adaptor, θ_i is an initial threshold set for the robot. As above, situational component weights, σ_k, ranging from 0 to infinity, represent the relative effect of each component. The additive factor τ is derived from user input. Finally, the weight for situation j, a_j, is a scaling factor ranging from 0 to 1 and is related to the necessity of an empathic response for a given situation.

The action-selection mechanism must address three interrelated problems. The foremost of these is the problem of how the emotional state should be accrued by the system. A moral emotion (e.g., shame, embarrassment, empathy, guilt), however, does not typically exist in a binary manner; rather it is present in variable amounts. Thus, it is also necessary to determine how much of said emotion should result from the situation and its resulting action. This should take into account that the same situation (e.g., playing a game with one's child) can evoke different moral emotions at different times (e.g., depending on the parent's level of patience that day, mood, etc.). Finally, it is not enough for the robot to merely accrue these models of moral emotions from the human participants. It is also necessary to define how the architecture interacts with the underlying behavioral system in order to express its moral support in some manner through behavioral change. In this case, it results in the selection of an ethical action from a suite of ethical frameworks which may or may not be producing conflicting choices. Our initial focus is on positive moral emotions. The aim is to maintain a partial theory-of-mind of the robot's human counterparts in the relationship in order to act in manner that fosters their emotional state in a manner consistent with enhancing their dignity. We also focus on reducing guilt, shame, and frustration that humans may experience while interacting with a robot.

In contrast to previous approaches to creating ethical autonomous systems, the intent of this research is to allow an autonomous system to adaptively utilize different ethics frameworks depending on the context and on its past experiences. We believe that this added flexibility results in more robust and ethical decision-making by the system. Figure 1 offers a high-level architecture of the system. Rather than attempting to implement each ethical framework directly (Utilitarianism, Kantianism, or Ross's moral duties), we create a collection of experiences capturing the context and each framework's action selection recommendation when faced with an ethically complex situation relating to two scenarios: game playing and pill sorting. Depending on one's perspective, these experiences can be viewed as cases for a case-based reasoning system, instances of data related to ethical situations, or a form of computational memories. Cases have been used to create ethical systems in the past [29].

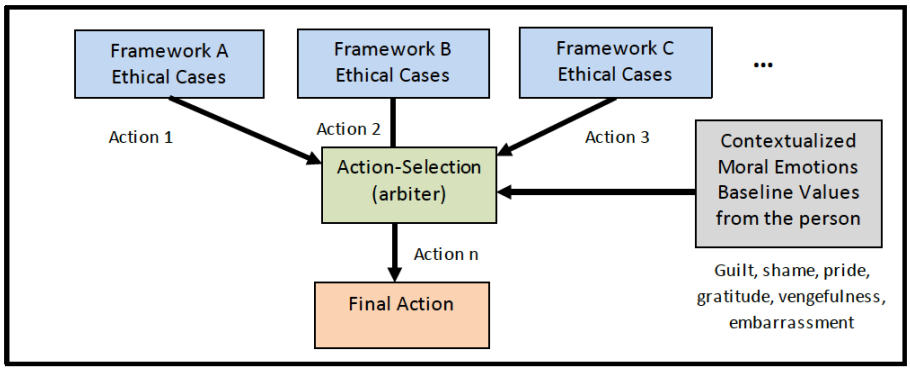

Figure 1. High level architecture. Multiple potentially contradictory ethical sources suggest actions to the action selection mechanism (arbiter) which is biased by the ongoing situation and the moral emotional state of the individuals in question.

The cases are represented in a computationally straightforward fashion. We have extensively used case-based reasoning in the past and now adapt methodologies used in the past (e.g.,

[30,31]). Case-based reasoning (CBR) is a very general Artificial Intelligence methodology for automatic synthesis of plans based on the observation that problem solvers often reuse an existing plan to solve a new problem if they have a plan that had worked for a similar problem in the past. CBR can enable artificial cognitive systems to assess context and situation, recall a relevant resolution to an ethical problem from the past (complete with explanatory information), and then provide an analogical mapping onto the current situation derived from a range of similarity measures, finally guiding the resulting action of the robot in a manner consistent with past beliefs regarding ethical behavior.

These ethical cases serve as a knowledge base capturing contextual information and the resulting decisions that were made. These cases are created by presenting people with the same ethically complex situations that the robot will face in each of the scenarios. Just as the Trolley Problem has been used to provide insight on a person's ethical decision-making process in relation to autonomous vehicles, [32] we develop a series of game playing and pill sorting problems to understand human decision-making in these scenarios. In this upcoming research, we will present the situations, using survey instruments, to a broad human subject population of approximately 100 to 200 people. This population will not directly be asked to use one of the three ethical frameworks; they will serve as the basis for the "folk" morality used to inform a robot's decisions. For comparison the same scenarios will be reviewed by ethics experts (ethics instructors for example). These experts will be asked to derive an action recommendation from each of the three frameworks. Initially, these situations are generated as textual descriptions.

6. Summary

This paper describes the motivation, background, and approach for an architectural framework that will select between ethical outcomes for a variety of situations generated by widely differing ethical frameworks: Kantianism, Utilitarianism, and Ross's moral duties. The action-selection process will be mediated by moral emotions to account for the affective state of the agent and the context in which it resides. Results from interaction with the general population in assessing which is the correct action to undertake for a specific circumstance, and ethics experts will be used as a baseline for this NSF-funded research project.

Acknowledgements

This research is funded by the National Science Foundation as part of the Smart and Autonomous Systems program under Grants No. 1849068 and 1848974.

References

1. Anderson, M., and Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, *28*(4), 15.
2 Bello, P., & Bringsjord, S. (2013). On how to build a moral machine. *Topoi*, *32*(2), 251-266.
3. Blass, J.A., and Forbus, K.D., (2015). Moral Decision-Making by Analogy: Generalizations versus Exemplars. In *AAAI*, pp. 501-507.
4. Sharkey, N. (2008). The Ethical Frontiers of Robotics, *Science*, (322): 1800-1801.
5. Abel, D., MacGlashan, J., & Littman, M. L. (2016). Reinforcement learning as a framework for ethical decision making. *Workshop 13 AAAI Conference on Artificial Intelligence*.
6. Scheutz, M., Malle, B., & Briggs, G. (2015). Towards morally sensitive action selection for autonomous social robots. In *Robot and human interactive communication, 2015 24th IEEE international symposium on* (pp. 492-497). IEEE.
7. Anderson, M., and Anderson, S. L. (Eds.). (2011). *Machine ethics*. Cambridge Univ. Press.
8. Iba, W., & Langley, P. (2011). Exploring moral reasoning in a cognitive architecture. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 33, No. 33.

9. Beavers, A. F. (2009). Between angels and animals: The question of robot ethics, or is Kantian moral agency desirable. In *Association for practical and professional ethics, eighteenth annual meeting, Cincinnati, Ohio, March* (pp. 5-8).
10. Bringsjord, S.; Arkoudas, K.; and Bello, P. (2006). Toward a General Logicist Methodology for Engineering Ethically Correct Robots. *IEEE Intel. Syst.* 21(4): 38– 44.
11. Saptawijaya, A., & Pereira, L. M. (2015). The potential of logic programming as a computational tool to model morality. In *A Construction Manual for Robots' Ethical Systems* (pp. 169-210). Springer, Cham.
12. Anderson, M., Anderson, S. L., & Berenz, V. (2019). A value-driven eldercare robot: Virtual and physical instantiations of a case-supported principle-based behavior paradigm. *Proceedings of the IEEE*, *107*(3), 526-540.
13. Johnstone, J. (2007). Technology as empowerment: A capability approach to computer ethics. *Ethics and Information Technolog,y* 9 (1): 73-87.
14. Sinnott-Armstrong, W., (2015). Consequentialism, *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.).
15. Alexander, L. and Moore, M., "Deontological Ethics", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.).
16. Ross, D., (1930*). The Right and the Good.* Oxford: Oxford University Press.
17. O'Fallon, M.J., and K. D. Butterfield. (2005). "A Review of the Empirical Ethical Decision-Making Literature: 1996-2003." *Journal of Business Ethics* 59, no. 4: 375-413.
18. Timmermann, Jens. (2013). Kantian Dilemmas? Moral Conflict in Kant's Ethical Theory. *Archiv für Geschichte der Philosophi,e* 95 (1):36-64.
19. Rachels, J. and S. Rachels, (2015). *The Elements of Moral Philosophy* (8th ed.).
20. Arkin, R.C., (2005). Moving Up the Food Chain: Motivation and Emotion in Behavior-based Robots, in *Who Needs Emotions: The Brain Meets the Robot*, Eds. J. Fellous and M. Arbib, Oxford University Press.
21. De Melo, C., Zheng, L. and Gratch, J., (2009). Expression of Moral Emotions in Cooperating Agents. *9th International Conference on Intelligent Virtual Agents.*
22. Arkin, R.C. and Ulam, P., (2009). An Ethical Adaptor: Behavioral Modification Derived from Moral Emotions, *IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA-09)*, Daejeon, KR.
23. Haidt, J. (2003). The Moral Emotions, in *Handbook of Affective Sciences*, Oxford Press.
24. Tangney, J., Stuewig, J., and Mashek, D., (2007). Moral Emotions and Moral Behavior, *Annu. Rev. Psychol.,* Vol. 58, pp. 345-372.
25. Smits, D., and De Boeck, P., (2003). A Componential IRT Model for Guilt, *Multivariate Behavioral Research*, Vol. 38, No. 2, pp. 161-188.
26. Cameron, C., Lindquist, K., and Gray, K. (2015). A Constructionist Review of Morality and Emotions: No Evidence for Specific Links Between Moral Content and Discrete Emotions**,** *Personality and Social Psychology Review* 2015, Vol. 19(4) 371–394.
27. Allen, C., Wallach, W., and Smit, I., (2006). Why Machine Ethics? *IEEE Intelligent Syst*ems, July.
28. Gazzaniga, M., (2005). *The Ethical Brain*, Dana Press.
29. McLaren, B. M. (2003). Extensionally Defining Principles and Cases in Ethics: An AI Model. *Artificial Intelligence Journal,* 150(1– 2): 145–1813.
30. Kira, Z. and Arkin, R.C., (2004). Forgetting Bad Behavior: Memory Management for Case-based Navigation, *Proc. IROS-2004*, Sendai, JP.
31. Ram, A., Arkin, R.C., Moorman, K., and Clark, R.J., (1997). Case-based Reactive Navigation: A case-based method for on-line selection and adaptation of reactive control parameters in autonomous robotic systems, *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 27, Part B, No. 3, pp. 376-394.
32. Wilson, J. R., & Scheutz, M. (2015). A model of empathy to shape trolley problem moral judgements. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on* (pp. 112-118). IEEE.