Experimental Methods for Moral Behaviour Analysis in Human-Robot Interaction

Francesco Perrone

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Computing Science
College of Science and Engineering
University of Glasgow

February 2023

This work has been partially supported by the UK Engineering and Physical Sciences Research Council (EPSRC) through the grant "*Socially Competent Robots*" (EP/N035305/1).

Abstract

Abstract text goes here.

Contents

List of Tables

List of Figures

Acknowledgements

Acknowledgements text goes here.

Declaration

With the exception of chapters 1, 2 and 3, which contain introductory material, all work in this thesis was carried out by the author unless otherwise explicitly stated.

1. Introduction

Moral decision making, is the cognitive process of choosing between competing moral judgments *i.e.*, mutually exclusive evaluations we make on what is right or wrong, good or bad, and that we use as motive, purpose and direction for our conscious, and practical behaviour.

a) **Cognitive Process** This term refers to the mental actions or operations that individuals use to acquire knowledge and understanding. It includes processes such as perception, memory, reasoning, decision-making, and problem-solving. Cognitive processes are essential for interpreting and interacting with the world;

b) **Behaviours:** In academic terms, behaviours are the observable actions or reactions of an individual in response to external or internal stimuli. These actions can be voluntary or involuntary and are influenced by various factors, including cognitive processes, emotions, and environmental conditions.

Moral decision making is the intricate cognitive process of choosing between competing moral judgments; these are mutually exclusive evaluations we make regarding what is right or wrong, good or bad. These judgments serve as the motive, purpose, and direction for our conscious and practical behaviour. This process involves an array of cognitive functions such as perception, memory, reasoning, and problem-solving, which collectively inform our moral evaluations and decisions. Moreover, these cognitive processes translate into behaviours, which are the observable manifestations of our moral choices. These behaviours, whether conscious or subconscious, reflect our internal moral deliberations and are influenced by a complex interplay of cognitive functions, emotions, and contextual factors. Hence, moral decision making encompasses both the mental operations that guide our judgments and the resultant actions that embody our moral principles in the practical realm.

The perception of direct gaze, that is, of other individual gaze directed at the observer, is know to influence a wide range of cognitive processes and behaviours.

1.1 Machines' Ethics

Machine Ethics is the subfield of Computer Science that develops methods and theories aimed at enabling machines to interact morally with their users in real-world scenarios. The role of Machine Ethics has received increased attention across a number of academic disciplines, in the past few years

[1].

A central reason for this encouraging circumstance is an unprecedented interdisciplinarity: researchers in Machine Ethics are now capable of freely drawing on scientific resources from well beyond the confines of their fields, a scientifically robust data that can now be integrated and used as a laboratory to verify e generalise more qualitatively philosophical outsets which were common of its foundational work [5, 6].

The broad concept of "artificial intelligence" (AI) encapsulates any form of synthetic computational mechanism that exhibits intelligent actions, which are complicated actions conducive to achieving objectives. We aim to refrain from confining "intelligence" strictly to tasks requiring human intellect, contrary to Minsky's proposal [25]. Thus, we include a wide array of machines, encompassing "technical AI" systems that demonstrate only limited learning or reasoning skills but excel in task automation, and "general AI" systems designed to establish a universally intelligent agent. AI tends to intertwine more with our existence than other technologies, hence the emergence of the "philosophy of AI". Possibly, this arises from the AI's endeavour to fabricate machines that possess attributes that we humans perceive as vital to our identity, such as the ability to feel, think, and show intelligence. The primary roles of an AI agent likely involve sensing, modelling, planning, and execution, but current applications extend to perception, text scrutiny, natural language processing (NLP), logical deduction, gameplaying, decision-making aids, data analysis, predictive analytics, along with self-operating vehicles and other robotic manifestations [34].

AI might employ various computational strategies to achieve these goals, like classic symbol-manipulating AI, cognitive inspired processes, or machine learning through neural networks [20, 33]. It's important to acknowledge that historically, the term "AI" was used as previously mentioned roughly between 1950-1975, followed by a period of skepticism during the "AI winter", approximately from 1975-1995, and was subsequently constrained. Consequently, areas like "machine learning", "natural language processing", and "data science" were typically not categorized as "AI". Around 2010, the usage expanded again, with at times nearly all of computer science and even high-tech being consolidated under "AI". Presently, it has transformed into a prestigious moniker, a thriving sector with substantial capital investment [32], and is on the brink of resurging hype. As Erik Brynjolfsson pointed out, it might empower us to virtually eliminate global poverty, massively reduce disease, and provide superior education to almost every person on earth [2].

While AI can solely be software-based, **robots are tangible machines capable of movement**. Robots are subject to physical effects, primarily via "sensors", and exert physical force onto the environment, typically through "actuators", such as a gripper or a rotating wheel. Therefore, autonomous vehicles or aircrafts are robots, and only a tiny fraction of robots are "Humanoid" (human-resembling), as depicted in films. Some robots employ AI, while others do not: Standard industrial robots rigidly adhere to fully defined scripts with minimal sensory input and devoid of learning or reasoning (approximately 500,000 such new industrial robots are deployed each year [23]). It is likely appropriate to state that although robotic systems incite more apprehension among the public, AI systems are more likely to significantly influence humanity. Moreover, AI or

robotic systems designed for a narrow range of tasks are less likely to pose new challenges than more flexible and independent systems. Hence, robotics and AI can be visualized as encompassing two intersecting categories of systems: those that are solely AI, those that are strictly robotic, and those that are a combination of both. Our interest spans all three; the focus of this article encompasses not just the intersection, but the amalgamation, of both categories. In the rapidly progressing domains of artificial intelligence (AI) and social robotics, the necessity of ethical deliberation and moral agency is paramount. As these technologies become increasingly sophisticated and entrenched in our everyday lives, timeless philosophical queries concerning purpose, potentiality, and morality gain renewed relevance. Ancient Greek philosophers endeavoured to delineate and comprehend human moral agency, a task that now confronts us in the context of AI and robotics. Drawing on the profound insights of philosophers like Aristotle, we can navigate and address the unique ethical conundrums raised by these technologies. However, it is crucial to recognise a prevalent shortcoming in the discourse on AI and robotics. Academics and authors in the field frequently employ terms such as "moral and morality", "ethics", "intentionality and agency", yet these concepts often lack a deep philosophical grounding [26]. This absence of philosophical understanding can lead to misconceptions and flawed assumptions, particularly in a field as nuanced as AI [10]. For instance, the application of "moral agency" to AI systems can be contentious, given that traditional interpretations of the term presuppose qualities like consciousness and intentionality that machines do not possess [17]. Similarly, there can be a tendency to anthropomorphise AI systems when discussing their 'ethics,' which can obfuscate the fact that their 'ethical' behaviours are entirely human-programmed [41]. In this paper, we strive not only to draw insightful parallels between ancient philosophy and contemporary ethical discussions in AI and social robotics but also to illuminate and correct potential misconceptions caused by a lack of philosophical understanding. By grounding our discussions in solid philosophical foundations, we hope to foster a more nuanced, accurate, and productive discourse on AI ethics.

Aristotle's teleological view of existence, as detailed in his collective works [7], interprets the universe as inherently intentional. He advocates that potentiality is in service of actuality, asserting that matter's essence lies in the prospect of adopting form[41], paralleling how an organism is endowed with sight for the purpose of perception. In this vein, every entity bears unique potentialities that spring from its form. Drawing upon this, a serpent, due to its form, possesses the capacity to undulate, implying it's naturally inclined towards this movement. The fulfilment of potential is directly tied to the realisation of its intended purpose.

This teleological paradigm serves as the foundation of Aristotle's ethical philosophy [35]. The form of humans confers upon them certain abilities. Hence, their purpose is intertwined with the proficient and complete utilisation of these capacities.

Transitioning to computational morality and robotics, Aristotle's teleological framework presents a compelling lens for analysis. Analogously, robots, initially devoid of purpose, derive their purpose from their programmed tasks and abilities. In a manner similar to Aristotle's view of matter waiting to receive form, a raw computational canvas exists to embrace coding and programming[40]. Mir-

roring an organism's sight intended for seeing, a robot is equipped with sensors designed to interact with its environment [31].

Each robot, through its specific programming or "form," carries certain capabilities. For instance, an autonomous vehicle, due to its form, has the ability to navigate, implying that it is programmed to do so. The extent to which a robot actualises its potential mirrors the success it achieves in fulfilling its designed purpose.

When Aristotle's teleological worldview is applied to computational morality in AI systems, it generates intriguing considerations. AI systems, due to their 'form' or programming, are vested with certain abilities, such as learning, analysing, and decision-making based on intricate algorithms [**?**]. Therefore, their 'purpose' can be seen as the maximal and effective application of these abilities, aiming to reach ethical decisions that align with their programmed ethical framework [1].

Aristotle's teleological views weren't formed in a vacuum, and they can be further contextualised within the larger discourse among Ancient Greek philosophers. For instance, Plato, Aristotle's mentor, maintained a theory of forms, emphasising an immaterial world of 'perfect' forms separate from our everyday world. Yet, Aristotle rejected this dualism, proposing instead that forms existed in objects and, crucially, it was this form that gave objects their purpose.

Aristotle's emphasis on the form and potentiality of a being can be intriguingly juxtaposed with the concept of "Levels of Abstraction" (LoA) proposed by Luciano Floridi [19]. Floridi suggests that understanding a system requires viewing it at the appropriate LoA, a conceptual lens that filters out unnecessary details and focuses on the information needed to understand or interact with the system. In computational terms, the 'form' of an AI system would correspond to its designed LoA. Just as Aristotle sees a being's form as key to understanding its purpose and potentiality, Floridi sees an AI's LoA as critical to understanding its function and capabilities. This highlights the parallels between ancient philosophical thought and contemporary information philosophy. This connection further emphasises the relevance of Aristotle's teleology to computational morality.

If we take the AI's designed LoA as its 'form', then the purpose of the AI system becomes fulfilling the functions and potentialities set out at this level. This mirrors the Aristotelian notion that an entity's purpose is tied to fulfilling its potentialities as dictated by its form. A complete understanding of computational morality, therefore, requires an appreciation of the designed LoA of the AI system. Just as Aristotle advocated for a nuanced understanding of an entity's form, so too does Floridi's framework encourage us to consider the appropriate LoA when grappling with moral issues in AI and robotics.

Aristotle serves as a starting point for this exploration due to his pivotal role in laying the groundwork of Western philosophical thought. His concept of teleology, or the purposefulness of all things and actions, has significantly influenced subsequent understandings of ethics and morality.

Moreover, his views on Actuality and Potentiality provide a useful lens through which to consider the capabilities and purpose of artificial intelligence. Nevertheless, it is crucial to appreciate that Aristotle's perspective is only the first of

many that we will engage with in this investigation. As we traverse the historical landscape of philosophical thought on morality and ethics, we will encounter a rich tapestry of ideas that each contribute uniquely to our modern grappling with these concepts in the context of AI and social robotics.

Within the realm of formal logic, the precision of definitions constitutes a bedrock. For instance, the rigorous delineation of a proposition as a statement with a definitive truth value - either true or false, but never both nor neither underpins all ensuing discourse. Logical connectives, such as 'and', 'or', and 'not', gain their operational power from the meticulously prescribed relationships they signify between propositions. The process of formulating complex logical rules and inferences becomes an orchestrated composition, owing its harmony to the preciseness of these core definitions [24]. In mathematics, the emphasis on defining primitive entities is equally profound. For example, in set theory, which provides a foundation for virtually all of mathematics, the concept of a set is primitive and left undefined. Instead, the properties and operations of sets are described by axioms, such as those proposed by Zermelo and Fraenkel [37]. In number theory, the definition of what constitutes a number has evolved over time, from the natural numbers to the inclusion of zero, negative numbers, rational numbers, real numbers, and complex numbers, each expansion necessitating a precise definition to avoid ambiguity and contradiction [30]. The rigorous defining of terms is far from a simple formality; it facilitates clear communication, reduces ambiguity, and enhances the richness of academic discourse. The vast terrain of interdisciplinary fields like AI and Social Robotics demands a similar level of precision and clarity in the definitions of often philosophically loaded terms like 'morality', 'ethics', and 'agency', especially given their diverse interpretations across various contexts [26].

Notes

[1]A search for the keyword '*Computational Morality*' alone on Google Scholar yielded an astonishing number of more than 39,000 results as of October 2021. However, as of today, this figure has significantly grown to about 86,200 results, indicating a substantial increase in literature on the subject over the past year. Furthermore, a search for the keyword '*Machine Ethics*' on Google Scholar produced an already staggering number of approximately 3,000,000 results as of October 2021. However, the figure has seen a remarkable growth, now standing at about 3,230,000 results, emphasising the continued expansion of research and scholarly engagement with the ethical aspects of artificial intelligence. These notable increases and changes in the figures for both 'Computational Morality' and 'Machine Ethics' highlight the growing prominence and visibility of these fields within the academic community. They signify the escalating interest among researchers, scholars, and ethicists in investigating the ethical dimensions of computational systems and the moral implications of their actions *at the least*. The significant growth in literature not only reflects a broader understanding of the ethical challenges posed by advancing technologies but also underscores the pressing need to address and discuss the ethical considerations associated with the design, deployment, and impact of computational systems in our society. It is worth noting that the figures provided here are based on a search conducted on Google Scholar as of July 9, 2024. Due to the dynamic nature of online databases, the exact figures may vary over time. Nonetheless, the substantial increase in publications on computational morality and machine ethics signifies the continuous expansion and significance of these fields in the realm of ethical inquiry. The rapid growth of research in the field of computational morality and machine ethics highlights its paramount importance in our increasingly technologically-driven world. As computational systems and artificial intelligence become more

integrated into various aspects of society, it is crucial to explore the ethical implications of their actions [**we are not doing this here**]. Understanding and addressing the moral dimensions of these systems is vital to ensure their responsible development, deployment, and impact on individuals and communities. The remarkable expansion of literature in computational morality is a testament to the urgency and significance of this research area. In fact, the rate of growth in this field often surpasses that of many other scientific and computer science-related disciplines, illustrating the heightened attention and recognition it receives. This exponential rise underscores the interdisciplinary nature of computational morality, drawing insights from philosophy, computer science, sociology, and other fields. It highlights the recognition among scholars, researchers, and practitioners that the ethical considerations and social implications of computational systems are integral to the advancement of technology and the well-being of society as a whole. By delving into computational morality, we pave the way for a future in which ethical principles guide the design, implementation, and use of intelligent systems, ensuring that they align with human values and promote the greater good.

2. On Moral Decision Making

But one thing is the thought, another thing is the deed, and another thing is the idea of the deed. The wheel of causality doth not roll between them.

Friedrich Nietzsche, *Thus Spoke Zarathustra* (1883)

Analysing the concept of *Moral Decision Making* in the context of predicate logic involves interpreting various linguistic elements within a logical framework.

- **The Word "Decision"**: In predicate logic, "Decision" can be a constant or a variable.
  - As a constant (for a specific decision), it might be represented as d.
  - As a variable (representing any decision), it could be denoted as x, where x is a decision.
- **The Noun Phrase "Decision Making"**: "Decision Making" can be interpreted as a function in predicate logic.
  - The function DecisionMaking(x) represents the output or consequence of making decision x.
- **The Adjective "Moral" in "Moral Decision Making"**: "Moral" is a modifier and can be viewed as a predicate.
  - The predicate Moral(DecisionMaking(x)) indicates that the decision-making process of x is of a moral nature.

A typical formula connecting these elements might be:

$$\forall x(\text{Decision}(x) \rightarrow \text{Moral}(\text{DecisionMaking}(x)))$$

This formula can be interpreted as: "For all x, if x is a decision, then the decision-making process of x is moral." It employs a universal quantifier ($\forall$) to express a general statement about all decisions.

In moral philosophy, these logical structures assist in defining and debating ethical theories and concepts, enabling a rigorous analysis of the nuances of moral decision-making.

The concept of *Moral Decision Making* can be more accurately represented in predicate logic by considering that not all decisions are inherently moral, but rather, they become moral under certain conditions.

Consider the revised approach:

- **Existential Quantification and Conditionality**: The formula should reflect that only some decisions fall under the category of moral decisions, contingent upon specific conditions being met.

A more realistic formula would be:

$$\exists x(C(x) \rightarrow (\text{Decision}(x) \wedge \text{Moral}(\text{DecisionMaking}(x))))$$

Here, $C(x)$ represents the specific conditions under which a decision x can be considered moral. The formula is interpreted as: "There exists some decision x such that if the conditions $C(x)$ are met, then x is a decision and the decision-making process concerning x is moral."

This formula acknowledges that morality in decision-making is not a universal attribute of all decisions, but rather a characteristic of certain decisions under specific circumstances. Identifying and analyzing these conditions $C(x)$ is a key aspect of ethical philosophy and moral reasoning.

I want to precisely narrow down the meaning of the word *Decision*, in the

In the discourse regarding the evolution of moral theories across philosophy and modern psychology, there emerges a nuanced interconnection where *emotional* and *rational* elements not only diverge but also integrate. This interconnection reveals the intricate complexities inherent in the formulation of moral models, transcending a mere dichotomy to embrace a more holistic, undefined perspective.

Definition 1 (Rational Model) *Moral decision making, is a cognitive process of choosing between competing moral judgments-* i.e.*, mutually exclusive evaluations we make on what is right or wrong, good or bad, and that we use as motive, purpose and direction for our conscious, practical behaviour.*

Definition 2 (Emotional Model) *Moral decision-making is an emotive process, wherein individuals navigate and choose between competing moral judgments* i.e.*, mutually exclusive evaluations we make on what is right or wrong, good or bad. This process is driven by emotional responses and intuitions, which guide and inform our conscious and practical behaviour, often preceding and shaping cognitive deliberation.*

Moral decision-making represents a cognitive exercise in the calculus of ethics. Within this framework of moral calculus, contemporary and classical scholars offer a spectrum of perspectives on the central role of emotional and cognitive faculties alike that incorporates both emotional and cognitive faculties.

It is worth delineating the concept of 'moral calculus' as distinct from *hedonistic calculus*. While the latter term typically refers to Benthamite utility maximization, often quantified in terms of pleasure and pain, 'moral calculus' serves as a broader framework for ethical deliberation. Unlike hedonistic calculus, which is generally rooted in consequentialist traditions, moral calculus navigates the complexities of diverse ethical systems, be they deontological, virtue-based, or others.

The philosophical canon profoundly integrates the conceptual distinction between

emotion-driven and reason-driven moral philosophies. This differentiation has seen considerable evolution over centuries, leading to a significant impact on contemporary psychological thinking, especially in the sphere of moral psychology. The nuanced separation of emotion-driven and reason-driven moral frameworks, deeply rooted in philosophical discourse, has evolved extensively over time, culminating in its marked influence on modern psychological studies, with a particular focus on moral psychology. This journey begins with the foundational works of ancient philosophy. In this era, thinkers like Plato in his 'Republic' delineate a clear preference for reason over emotion in guiding ethical conduct. Aristotle, in his 'Nicomachean Ethics,' echoes this sentiment to some extent by emphasizing rational virtues, yet he also acknowledges the significant role emotions play in ethical existence.

Moving forward into the Enlightenment, this conceptual distinction was further crystallized. A paradigmatic figure of this era, Immanuel Kant, championed a morality firmly rooted in reason and universal maxims in his works, such as 'Critique of Pure Reason' and 'Groundwork for the Metaphysics of Morals,' thereby relegating emotions to a subsidiary role. This period marked a significant shift toward a rationalist perspective in moral philosophy.

However, this shift was met with a counterpoint in the British empirical tradition. Figures like David Hume presented a challenge to the Kantian rationalism. In his 'A Treatise of Human Nature,' Hume provocatively posited that reason is subordinate to passions, thereby anchoring moral judgments in emotional responses. This perspective from British empiricists highlighted the importance of emotions, or 'sentiments', in moral considerations, offering a contrasting view to the prevailing rationalist approach.

Moral decision making, is a cognitive process of choosing between competing moral judgments- *i.e.*, mutually exclusive evaluations we make on what is right or wrong, good or bad, and that we use as motive, purpose and direction for our conscious, practical behaviour.

Moral decision-making is primarily an emotive process, wherein individuals navigate and choose between competing moral judgments *i.e.*, mutually exclusive evaluations we make on what is right or wrong, good or bad. This process is driven by emotional responses and intuitions, which guide and inform our conscious and practical behaviour, often preceding and shaping cognitive deliberation.

Hence, in the discourse of moral calculus, certain schools of thought, notably those propounded by Haidt (2012) and Greene (2007), assert with compelling vigour that the substratum of emotional faculties, rather than those of the cognitive domain, governs the architecture of ethical decision-making. This viewpoint finds a harmonic resonance in the philosophical canon, corroborated by seminal treatises such as Nussbaum's 'Upheavals of Thought' (2001) and Damasio's 'Descartes' Error' (1994).

The theoretical edifice of moral calculus, while intellectually robust, gains tangible relevance when juxtaposed with empirical and phenomenological data. Bridging these domains allows for a more encompassing understanding of moral decision-making, marrying the abstract with *the* concrete, *the* theoretical with *the* expe-

riential.

For the empirical aspect:

Neuroscientific research offers valuable empirical insight into the machinery of moral cognition. Studies have implicated regions like the prefrontal cortex and the amygdala in the ethical decision-making process (Greene et al., 2001; Decety & Cacioppo, 2012). These findings suggest that our 'moral calculus' may indeed have a tangible neurological substrate, grounding ethical theory in the biological realm.

For the phenomenological aspect:

Complementing these empirical observations, phenomenological accounts provide a subjective lens through which moral decision-making can be examined. Authors such as Sartre and Merleau-Ponty have explored the existential dimensions of choice, capturing the lived experience of moral deliberation (Sartre, 1943; Merleau-Ponty, 1945).

These works serve to enrich our understanding of 'moral calculus' by infusing it with the subjective quality of human experience.

Emotion-Centered Models: Some theories argue [**?**] that emotional processes, rather than cognitive ones, are at the core of moral decision making. Your definition may not adequately capture the emotive factors often considered essential. onathan Haidt's work in "The Righteous Mind" explores the role of emotional intuition in moral judgments, arguing that reasoning often follows, rather than guides, our moral intuitions [45]

Practical behaviour is a term widely used across philosophy and psychology, it's challenging to create an exhaustive chronological definition because the term does not correspond to a singular theory or concept that has evolved over time in a linear fashion. Instead, it has been interpreted and applied differently depending on the context, theoretical framework, or school of thought.

Practical behaviour in philosophy: has been interpreted in various ways across different philosophical schools of thought. 1) In Aristotelian philosophy, practical behaviour is associated with "praxis" or action guided by moral virtue aimed at the good life. Practical wisdom ("phronesis") is crucial here as it guides one's decisions and actions in accordance with moral virtue. 2) Immanuel Kant distinguished between theoretical reason (used to understand the natural world) and practical reason (used to govern behaviour and moral decision-making). For Kant, practical behaviour is guided by the categorical imperative, an absolute moral law. 3) In the late 19th and early 20th century, the pragmatists (like William James and John Dewey) viewed practical behaviour as action informed by the effects that such behaviour would bring about.

Practical Behaviour in Psychology has been understood as observable actions and reactions to stimuli in behaviourism, deeply intertwined with internal cognitive processes during the cognitive revolution, and as a complex interplay of cognitive processes, emotional states, individual traits, and environmental influences in contemporary psychology. 1) In the behaviourist approach (Early 20th Century)

pioneered by John Watson and B.F. Skinner, practical behaviour is understood in terms of observable actions and reactions to stimuli, often studied through conditioning processes. 2) With the cognitive revolution (Mid-20th Century), practical behaviour started to be seen as deeply intertwined with internal cognitive processes like problem-solving, decision-making, and planning. 3) Social-Cognitive Theory (Late 20th Century): Albert Bandura's social-cognitive theory emphasised the role of observational learning, self-efficacy, and goal setting in practical behaviour. **Modern times**: today, in ethics and action theory, practical behaviour typically refers to behaviour guided by practical reason, that is, reason concerned with action and decision-making. This involves deliberation about means and ends, moral obligations, and the values at stake in different courses of action. Similarly in Contemporary Psychology, practical behaviour is understood as a complex interplay of cognitive processes, emotional states, individual traits, and environmental influences. It is typically studied in context-specific terms, such as health behaviour, consumer behaviour, or prosocial behaviour.

This is quite an encompassing scope, but there are inevitable aspects of the broader discourse on moral decision making that we need to include in this purview.

— the following needs to be integrated in the text —

The term "practical behaviour" is a broad one, encompassing a wide range of actions that an individual might take in their everyday life. These can range from simple behaviours like brushing teeth or driving to work, to more complex ones like making a significant decision about one's career or personal life."Moral behaviour," on the other hand, is a subset of practical behaviour. It refers specifically to actions that involve moral or ethical considerations. In other words, all moral behaviours are practical behaviours, but not all practical behaviours are moral behaviours. The distinguishing feature is the presence of moral or ethical considerations in the motivations, implications, or consequences of the action. So, in response to your question, the key to understanding the difference between "practical behaviour" and "moral behaviour" does indeed lie in understanding the specific meaning and implications of "behaviour" in these contexts. However, it's also crucial to consider the specific nature and context of the action itself, including the intentions behind it and its potential consequences. the key to understanding the difference between "practical behaviour" and "moral behaviour" does indeed lie in understanding the specific meaning and implications of "behaviour" in these contexts. However, it's also crucial to consider the specific nature and context of the action itself, including the intentions behind it and its potential consequences. Moral domain: A behaviour typically falls within the moral domain when it pertains to questions of right and wrong, fairness, justice, harm, and welfare. So, for instance, deciding to donate to charity falls within the moral domain because it involves considerations about the welfare of others. Playing the piano, on the other hand, would generally fall outside the moral domain because it's largely a personal interest or skill, not directly associated with the welfare or rights of others. Intention and motive: Moral behaviour often involves a level of intentionality, where the individual acts with a specific purpose or motive that is morally charged. An individual who donates to charity with the motive of helping others is engaging in moral behaviour. In contrast, an

individual who plays the piano for personal enjoyment is engaging in a practical behaviour that isn't inherently moral or immoral. Consequences: The potential or actual impact of behaviour on others also plays a crucial role in determining its moral status. Behaviours with positive or negative impacts on others are often evaluated on a moral basis. **Additional notes**: 1) *Interaction of factors determining behaviour:* In both philosophy and psychology, behaviour is viewed as a result of a complex interplay of multiple factors, including cognitive processes, emotional states, individual traits, and environmental influences. 1) Cognitive processes: Cognitive processes, such as perception, memory, decision-making, and problem-solving, play a critical role in practical behaviour. For example, decision-making theories, such as the Dual Process Theory, suggest that people use both intuitive (automatic, fast, and emotional) and deliberative (slow, controlled, and logical) systems in guiding their behaviours [38]. 2) Emotional states: Emotions can also guide our behaviour. The James-Lange theory of emotion suggests that our emotional experiences are a response to our bodily reactions to a stimulus. For example, we don't run away because we're afraid; instead, we're afraid because we see ourselves running[39]. 3) Individual traits: Personality traits influence how individuals interpret and respond to their environment. The Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) have been linked to various behavioural outcomes. For instance, high levels of conscientiousness have been associated with better job and academic performance[13]. 4) Environmental influences: Social and physical environments shape behaviour. Social Cognitive Theory emphasises the reciprocal nature of this relationship: our behaviour can both influence and be influenced by our environment. For example, observational learning suggests we learn behaviours by observing others, while self-efficacy can determine how we respond to challenges[9]. Modern psychology acknowledges that many behaviours are driven by processes outside of conscious awareness. For instance, implicit bias research shows that we often harbour unconscious biases that can influence our behaviour, including decision-making and interpersonal interactions[21].Decision-making is not always rational and is often influenced by cognitive biases. For example, the 'availability heuristic' suggests that people are more likely to consider information that's easily retrievable when making decisions, which may not always lead to accurate or optimal outcomes[42].This interdisciplinary field combines psychology and economics to understand decision-making and behaviour. For example, the concept of 'nudge theory' suggests that subtle changes in how choices are presented can significantly influence decisions and behaviour, a principle that has been applied in various domains like healthcare, finance, and public policy[43]. These insights suggest that our understanding of practical behaviour needs to be multifaceted, taking into account not only conscious, deliberate processes but also unconscious influences and the way cognitive biases and heuristics shape our decisions and actions. They also underscore the importance of considering the individual within their social and environmental context

Moral decisions theories are often analysed into components features such as the model of judgment adopted- whether factual or normative, rational of affects laden- its causes, and the ethical outset it seems to follow. All three components happen to be useful for identifying and organise methods and work done in Computational Ethics since they are easily linked to different scientific ap-

proaches adopted in the field, and their basis deeply root into both philosophical and psychological theories which have deeply inspired and implicitly shaped the objectives set for this field in the past two decades.

In particular, most modern philosophers have frequently written about the conflict between factual and normative judgments [**?**], between reason and emotions [**?**], and between normative and motivating reasoning [**?**] three dichotomies.

2.0.1 Normative Non-Ethical agents

A moral decision is what we *judged* necessary to resolve conflicts with an explicitly moral dimension via special type of judgements which often called *normative* or *value judgements*: responses to stimuli with a moral dimension. Normative judgements assert or deny what *ought* to be the case whether or not it is *actually* the case (see figure 2.1, page 12), in contrast to factual judgements which assert or deny facts that *are the case* or a properly justified believe.

Factual judgements assert or deny facts that *are the case* or a properly justified believe, while normative judgements assert or deny what *ought* to be the case whether or not it *actually* is the case.

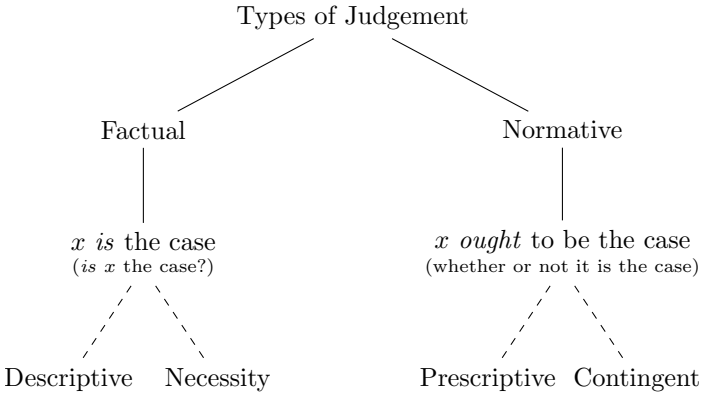

Figure 2.1: This distinction presuppose a sufficient prior understanding of the relevant uses of *is* and *ought* (or *should*) which will not discuss in details here. It is important to notice that, the presence of such marks as is neither a sufficient nor necessary criterion for the distinction we make, due to the striking variability of the relevant uses of the two words in every day language. For example, the sentence 'copper should be a metal' is not intended to be normative, and 'murder is evil' is not meant to be factual. Some philosophical theories claim that moral judgements lack of some desirable properties that factual statements have such as *objectivity* or *truth-apt*.

Judgements such as 'copper is a metal' [**?**] or '$2 + 2 = 4$' express what is the case because they are *truth-apt* judgements which means that they are either true or false in there being some corresponding *fact* which settles the question of their truth value. In simple terms, we cannot have different opinions on '$2 + 2 = 4$' because there exists a system of mathematical principles that, if accepted, makes us committed to the believe that '$2 + 2 = 4$': there are corresponding *facts* that make these locutions true or false.

On the other hand, judgements such as 'innocents ought not to be punished' [**?**] or 'wrongful killing is always wrong' are judgments about what *ought* to be the case but that do not have any corresponding fact that makes it true or false. Can machines grasp the difference between the two? In contrast with other more empirical judgments, moral judgements seem to have an intrinsic connection to motivation and action, for they form in us a uniquely bonding intentions to perform a behaviour, and motivate us to act in accordance with it [**?**].

Moor in [**?**] was one of the fist to examine how this distinction is relevant for a predominate class of works in Machine Ethics. Moor noticed that ordinary computers are designed with a purpose in mind, they are *normative agents* in the sense that they perform something on our behalf, executing rule-based instructions of which efficacy can be assessed assessed. However, ethical agents are those that perform actions with an ethical impact (positive or negative), but not by being constrained by their designers as this would not count has ethical act by the the very same definition of Ethics.

2.0.2 Other

Hence, genuine moral decisions must have as 'end-product', actions or inactions. In The Language of Morals [**?**], R.M. Hare, one of the leading British moral philosopher of the twentieth century, gives this clear characterisation:

> If we were to ask of a person "What are his moral principles?" the way in which we could be most sure of a true answer would be by studying what he did. He might, to be sure, profess in his conversation all sorts of principles, which in his actions he completely disregarded; but it would be when, knowing all the relevant facts of a situation, he was faced with choices or decisions between alternative courses of action, between alternative answers to the question "What shall I do?", that he would reveal in what principles of conduct he really believed. The reason why actions are in a peculiar way revelatory of moral principles is that the function of moral principles is to guide conduct. ([**?**])

By the same token, the main objective of Machine Ethics is to develop implicit ethical agents that is to say, machines that have been programmed in a way that can decide on actions with an ethical impact on their environment. Machine Ethics revolves around a precise subset of decision-type, since not all decisions have a moral dimension, and therefore not all types of judgments are relevant to morality. For example, whether I should get a frosty cold drink on a hot day using my last pound is not a moral decision. Whether I should use my last pound to get a cold drink, or give it to the women begging for money, appears to be. Both are instances of decision concerned with actions, they drive *goal-oriented behaviours* in which our perceptual and memory system support decisions that determine our *actions* [**?**].

2.1 Extended Types of Judgments

Typically, judgments are classified into two primary types: factual (descriptive or necessary) and normative (prescriptive or contingent). However, these categories, though fundamental, may not fully encompass the range of judgments humans engage with. Various disciplines, including philosophy, psychology, and mathematics, suggest other classifications or subcategories.

While factual and normative categories provide a foundational classification of judgments, the diversity and complexity of human thought suggest the utility of additional categories. The appropriateness of any specific set of categories, however, depends on the nature of the subject matter and the research questions at hand.

1. **Value Judgments:** Value judgments focus on the worth, importance, or intrinsic merit of a subject. As a subset of normative judgments, they often pertain to ethical or moral dimensions. However, their unique emphasis on 'value' might warrant a separate consideration.
2. **Aesthetic Judgments:** Aesthetic judgments concern beauty or other aesthetic attributes. Although they might be regarded as a form of value or normative judgments, the discipline of aesthetics often treats them as a distinct category due to their specialised focus.
3. **Prudential Judgments:** Prudential judgments, often used in economics, decision theory, and practical ethics, consider what is prudent or practically wise. These judgments typically involve an interplay of both descriptive and normative elements.
4. **Probabilistic Judgments:** Probabilistic judgments, prevalent in statistics, psychology, and decision theory, assess the likelihood or probability of a given event or condition. They often require a balance between empirical data and theoretical models.
5. **Counterfactual Judgments:** Counterfactual judgments, commonly used in philosophy and cognitive psychology, speculate on alternate realities or conditions. These judgments often hinge on the ability to imagine and reason about hypothetical situations.
6. **Analytic Judgments:** In Kantian philosophy, analytic judgments are those in which the predicate concept is included within the subject concept. These judgments are typically tautological and contrast with synthetic judgments.
7. **Synthetic Judgments:** Kant also proposed synthetic judgments, wherein the predicate concept is not contained within the subject concept. They can be classified further into a priori (based on reasoning independent of experience) and a posteriori (based on experience).

for example, the judgments made in physics, like those in other scientific disciplines, can be seen to fall into several categories depending on the specific context. Much of the work in physics involves making *descriptive (factual) judgments* about the nature of the physical world. These judgments are usually based on

observation and experimentation and aim to accurately describe how the world is. While less common in physics than in other fields such as ethics, *prescriptive (normative) judgments* are sometimes made in the context of methodological rules about how to do physics. Physics often involves making *probabilistic judgments*. In quantum mechanics, the behaviour of particles is often described in terms of probabilities rather than definite outcomes. Physicists also frequently make *counterfactual judgments*, considering what would happen under different hypothetical scenarios. The distinction between *analytic and synthetic judgments* is also relevant in physics. An example of an analytic judgment in physics might be a mathematical truth that holds by definition within a certain model, while a synthetic judgment might be a statement about the physical world that is supported by empirical evidence.

The judgments made in physics, like those in other scientific disciplines, can be seen to fall into several categories depending on the specific context. Drawing upon Chalmers' work on the philosophy of science [14], we find that much of the work in physics involves making *descriptive (factual) judgments* about the nature of the physical world. These judgments are usually based on observation and experimentation and aim to accurately describe how the world is. While less common in physics than in other fields such as ethics, *prescriptive (normative) judgments* are sometimes made in the context of methodological rules about how to do physics, a concept explored by Laudan in his work on normative naturalism [15]. Physics often involves making *probabilistic judgments*. In quantum mechanics, the behaviour of particles is often described in terms of probabilities rather than definite outcomes [12]. Physicists also frequently make *counterfactual judgments*, considering what would happen under different hypothetical scenarios, a concept explored in the work of Woodward [16]. The distinction between *analytic and synthetic judgments* is also relevant in physics. Drawing on Bird's exploration of Kuhn's philosophy [11], we see that an example of an analytic judgment in physics might be a mathematical truth that holds by definition within a certain model, while a synthetic judgment might be a statement about the physical world that is supported by empirical evidence. In practice, many judgments in physics may involve a combination or an interplay of these types. The specific context and objectives of the work play a large role in determining which types of judgments are most relevant.

In practice, the types of judgments made in physics often involve a mixture of these categories. For instance, a descriptive judgment about the behaviour of a particle might be based on a combination of observation (a synthetic judgment) and mathematical reasoning (often involving analytic judgments). Thus, the understanding and classification of judgments in physics, like in other fields, benefit from a nuanced approach.

In a field such as Computer Science, a discipline that often intersects with logic, mathematics, and engineering, several types of judgments can be identified. Much of the work in computer science involves making *descriptive (factual) judgments*. These often take the form of specifying the behaviour of algorithms or systems, such as a judgment about the time complexity of a particular sorting algorithm. *Prescriptive (normative) judgments* are also found in computer science, often relating to best practices for coding, architectural decisions in system design, or

ethical considerations in AI development. *Analytic judgments*, where the predicate is contained within the subject, often emerge from logical deductions that follow from the definition of a concept or operation. *Synthetic judgments*, which refer to empirical findings that don't just follow from definitions, might involve observations about the performance of certain algorithms in specific contexts. Especially in areas like machine learning and algorithm analysis, computer scientists often make *probabilistic judgments*, like assessing the probability of a certain outcome given a set of inputs. In troubleshooting, system design, or in planning the development process, *counterfactual judgments* often play a significant role as computer scientists consider alternate scenarios or possibilities.

The rise of fields such as AI ethics and Human-Computer Interaction (HCI) has brought attention to *value judgments* in computer science. These might concern what constitutes fair treatment in an algorithm's decision-making process, for example. In practice, many judgments in computer science may involve a combination or an interplay of these types. The specific context and objectives of the work play a large role in determining which types of judgments are most relevant.

2.2 A definition of judgment

So, what is *judgment*?

A *judgment* has been defined differently across various fields. From a logical and mathematical perspective, it carries specific interpretations. In formal logic, a judgment is typically understood as an assertion that a proposition is true. This idea can be represented as follows:

$$J(P)$$

Here, J denotes the judgment operation and P is a proposition. The entire expression, $J(P)$, is read as "*it is judged that* P".

In mathematics, a judgment can be considered akin to a function. If we think of a judgment as mapping from a set of premises to a conclusion, we can represent it in a similar way to a function:

$$J : P \to C$$

Here, J is the judgment, P represents the premises, and C is the conclusion. This can be understood as a judgment J mapping a set of premises P to a conclusion C. Note, however, that this is a rather abstract and non-standard interpretation. Judgments in mathematics and logic are more typically represented as statements or propositions that are asserted to be true. German logician Gottlob Frege's work in the field of logic provides valuable insight into the concept of judgment. His Begriffsschrift, or concept script, was a formal language of logic devised to represent clear, logical thoughts. In Frege's system, judgments about a proposition can be symbolically expressed and manipulated.

3. Experimental Methods

During the past decade, new emerging technologies have caused profound changes in the way we communicate and interact [27]. Some of these changes have affected certain aspects of human behaviour and caused psychiatric disorders [46]. These technologies have fundamentally altered how we connect with others, potentially exacerbating feelings of loneliness despite increased opportunities for connection. The role that modern technologies—such as mobile communications, digital interaction platforms, and interactive humanoid robots might play in shaping these dynamics is critical, influencing not only interpersonal communications but also moral decision-making in complex social settings [3, 4, 8, 18, 44]. Furthermore, technologies that increase interactive opportunities may not necessarily enhance the quality or *ethical dimensions* of those interactions, which are crucial in scenarios involving moral choices [51, 52, 53, 54]. The constant presence of interactive technologies can lead to a reshaping of social norms and behaviours, which might lead to more engaged or more detached human responses depending on the context and implementation [40, 41].

Foundational insights from studies such as [46] set the stage for a deeper exploration into how contemporary communication technologies, particularly humanoid robots, might amplify or mitigate these effects by altering the quality and nature of social interactions in both visible and subtle ways.

This work presents experiments based on the Watching Eye effect, the tendency of people to behave more honestly or more pro-socially when they have the impression of being observed. In particular, the experiments of this work show that the presence of a robot is associated to a lower tendency to donate to a charity despite the presence of a Watching Eye stimulus (the picture of a child portrayed on the brochure of a Non-Governmental Organization providing medical care in poor countries). The tendency to donate was measured in terms of actually donated money and the results show that people donate roughly one and half times as much when there are no robots (a statistically significant difference). This suggests that, while not necessarily being involved in moral decisions, robots can still be associated to changes in the way people (possibly users) make decisions involving a moral dimension.

The *Watching Eye* effect is the tendency of people to behave more honestly or more pro-socially when they feel observed [47], whether such a feeling results from the presence of pictures depicting eyes [48], from the belief in a supernatural being that can see everything [49, 50], or from any other factors. The goal of this article is to investigate the interplay between the Watching Eye effect and the presence of humanoid robots, a technology expected to play an increasingly more important role in everyday life. In particular, the experiments of this work show that there is an association between the presence of a robot and the observable consequences of the Watching Eye effect.

3.1 The Influence of Observational Presence on Human Behavior: Experimental Insights from Human-Robot Interactions

4. Methodology

Here is another chapter to explain how the work was carried out.

A. Derivation of the equation

This is such boring material that it has been relegated to an appendix. Let's check an equation:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{A.1}$$

Let's hope I got it correct.

Bibliography

[1] Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine Ethics*. Cambridge University Press.

[2] Anderson, J., Rainie, L., and Luchsinger, A. (2018). *Artificial Intelligence and the Future of Humans*. Pew Research Center.

[3] Allcott, Hunt, Braghieri, Luca, Eichmeyer, Sarah, and Gentzkow, Matthew (2020). *The welfare effects of social media*. American Economic Review, 110(3), 629-76.

[4] Auxier, Brooke, and Anderson, Monica (2021). *Social media use in 2021*. Pew Research Center.

[5] Allen, Colin and Wallach, Wendell and Smit, Iva. (2006). *Why machine ethics?*, In: IEEE Intelligent Systems, 21(4), pp. 12–17. IEEE.

[6] Allen, C., & Wallach, W. (2012). *Moral machines: contradiction in terms or abdication of human responsibility*. In *Robot ethics: The ethical and social implications of robotics* (pp. 55–68). MIT Press Cambridge. Mass.

[7] Aristotle. (1984). *The Complete Works of Aristotle: The Revised Oxford Translation*. Princeton University Press.

[8] Bail, Christopher A. (2021). *Breaking the social media prism: How to make our platforms less polarizing*. Princeton University Press.

[9] Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ, 1986.

[10] Bryson, J. J. (2010). *Robots should be slaves*. In Close engagements with artificial companions: Key social, psychological, ethical and design issues (pp. 63-74). John Benjamins Publishing.

[11] Bird, A. (2000). *Thomas Kuhn*. Princeton University Press.

[12] Bricmont, J. (2016). *Making Sense of Quantum Mechanics*. Springer.

[13] Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and individual differences*, 13(6), 653-665.

[14] Chalmers, A. F. (2013). *What is this thing called science?* Hackett Publishing.

[15] Laudan, L. (1987). Progress or Rationality? The Prospects for Normative Naturalism. *American Philosophical Quarterly, 24(1), 19-31.*

[16] Woodward, J. (2007). *Making things happen: A theory of causal explanation.* Oxford university press.

[17] Dennett, D. C. (1971). *Intentional systems*. The Journal of Philosophy, 68(4), 87-106.

[18] Dwyer, Ryan J., El-Bardicy, Mostafa, and Hakami, Tahani (2020). *Seeking and avoiding digital distractions in the workplace*. Information Systems Journal, 30(5), 845-874.

[19] Floridi, L. (2008). *Levels of Abstraction and the Foundation of Computational Ethics*. APA Newsletter on Philosophy and Computers, 8(1), 3-5.

[20] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

[21] Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California law review*, 94(4), 945-967.

[22] Hampton, K. N., Sessions, L. F., Her, E. J., and Rainie, L. (2009). *Social isolation and new technology*. Pew Internet and American Life Project.

[23] International Federation of Robotics (IFR). (2019). *World Robotics Report*. IFR.

[24] Mendelson, E. (2009). *Introduction to mathematical logic*. CRC Press.

[25] Minsky, M. (1985). *The Society of Mind*. Simon and Schuster.

[26] Moor, J. H. (2006). *The nature, importance, and difficulty of machine ethics*. IEEE intelligent systems, 21(4), 18-21.

[27] Pantic, I. (2014). *Online social networking and mental health*, Cyberpsychology, Behavior, and Social Networking, volume 17, number 10, Mary Ann Liebert Inc 140 Huguenot Street 3rd Floor New Rochelle NY 10801 USA.

[28] Pantic, Maja and Vinciarelli, Alessandro (2014), *Social signal processing*, The Oxford handbook of affective computing, page 84

[29] Primack, B. A., Shensa, A., Sidani, J. E., Whaite, E. O., Lin, L. Y., Rosen, D., Colditz, J. B., Radovic, A., and Miller, E. (2017). *Social media use and perceived social isolation among young adults in the U.S.*, American Journal of Preventive Medicine, 53(1), 1-8. DOI: 10.1016/j.amepre.2017.01.010

[30] Russell, B. (1919). *Introduction to Mathematical Philosophy*. London: George Allen & Unwin.

[31] Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Malaysia; Pearson Education Limited.

[32] Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J.C., Lyon, T., Etchemendy, J. (2018). *The AI Index 2018 Annual Report*. AI Index Steering Committee, Human-Centered AI Initiative, Stanford University.

[33] Silver, D. et al. (2018). *A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play*. Science, 362(6419), 1140-1144.

[34] Stone, P. et al. (2016). *Artificial Intelligence and Life in 2030.* One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University.

[35] Taylor, C. (1985). *Human Agency and Language: Philosophical Papers, Volume 1.* Cambridge University Press.

[36] Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other.* Basic books.

[37] Zermelo, E. (1908). *Investigations in the foundations of set theory I.* In From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Ewald, W. (ed.), Oxford University Press.

[38] Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

[39] James, W. (1884). What is an Emotion?. *Mind*, 9(34), 188-205.

[40] Misra, S., Cheng, L., Genevie, J., and Yuan, M. (2016). *The iPhone Effect: The Quality of In-Person Social Interactions in the Presence of Mobile Devices.* Environment and Behavior, 48(2), 275-298.

[41] Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other.* New York: Basic Books.

[42] Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.

[43] Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

[44] Vosoughi, Soroush, Roy, Deb, and Aral, Sinan (2018). *The spread of true and false news online.* Science, 359(6380), 1146-1151.

[45] Haidt, Jonathan (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion.* Pantheon

[46] Xerxa, Yllza and Rescorla, Leslie A and Shanahan, Lilly and Tiemeier, Henning and Copeland, William E., (2023) *Childhood loneliness as a specific risk factor for adult psychiatric disorders*, Psychological Medicine, Volume 53 number 1, pages 227–235, Cambridge University Press.

[47] Oda, R., Kato, Y., & Hiraishi, K. (2015). *The watching-eye effect on prosocial lying.* Evolutionary Psychology, 13(3), 1474704915594959. Los Angeles, CA: Sage Publications.

[48] Atran, S. & Norenzayan, A. (2004). *Religion's Evolutionary Landscape: Counterintuition, Commitment, Compassion, Communion.* Behavioral and Brain Sciences, 27(6), 713-770.

[49] Bering, J.M., McLeod, K., & Shackelford, T.K. (2005). *Reasoning about dead agents reveals possible adaptive trends.* Human Nature, 16(4), 360-381.

[50] Shariff, A.F. & Norenzayan, A. (2007). *God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game.* Psychological Science, 18(9), 803-809. Los Angeles, CA: SAGE Publications.

[51] Sharkey, A., & Sharkey, N. (2010). *The crying shame of robot nannies: an ethical appraisal*. Interaction Studies, 11(2), 161-190.

[52] Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford: Oxford University Press.

[53] Lin, P., Abney, K., & Bekey, G.A., eds. (2012). *Robot ethics: The ethical and social implications of robotics*. Cambridge, MA: MIT Press.

[54] Bryson, J.J. (2010). *Robots should be slaves*. In Close engagements with artificial companions: Key social, psychological, ethical and design issues (pp. 63-74). Amsterdam: John Benjamins Publishing Company.