Techné: Research in Philosophy and Technology ISSN: 1091-8264
23:3 (2019): 425–455 DOI: 10.5840/techne2019122109

Robotic Nudges for Moral Improvement through Stoic Practice[1]

Michał Klincewicz

Abstract: This article offers a theoretical framework that can be used to derive viable engineering strategies for the design and development of robots that can nudge people towards moral improvement. The framework relies on research in developmental psychology and insights from Stoic ethics. Stoicism recommends contemplative practices that over time help one develop dispositions to behave in ways that improve the functioning of mechanisms that are constitutive of moral cognition. Robots can nudge individuals towards these practices and can therefore help develop the dispositions to, for example, extend concern to others, avoid parochialism, etc.

Key words: moral improvement, moral development, robotic companions, social robots, nudges, Stoic ethics

1. Introduction

Robots that interact with humans are not typically designed to help improve moral cognition or moral behavior, broadly construed. Nonetheless, they could in principle serve as surrogates for moral teachers or compliment whatever moral education children get at home or school. One way that robots could do this is by being a source of nudges (Thaler and Sunstein 2008). In a recent review of the debate on nudges, Cass Sunstein defines them in the following comprehensive way:

> Nudges are interventions that steer people in particular directions but that also allow them to go their own way. A reminder is a nudge; so is a warning. A GPS nudges; a default rule nudges. To qualify as a nudge, an intervention must not impose significant material incentives. A subsidy is not a nudge; a

Michał W. Klincewicz, Tilburg University, Cognitive Science and Artificial Intelligence, Warandelaan 2, 5037 AB Tilburg, Netherlands; Jagiellonian University, Institute of Philosophy, Department of Cognitive Science, Grodzka 52, 31-048, Kraków, Poland; M.W.Klincewicz@uvt.nl.

> tax is not a nudge; a fine or a jail sentence is not a nudge. To count as such, a nudge must fully preserve freedom of choice. If an intervention imposes significant material costs on choosers, it might of course be justified, but it is not a nudge. Some nudges work because they inform people; other nudges work because they make certain choice easier; still other nudges work because of the power of inertia and procrastination. (Sunstein 2015, 7–8)

There are other definitions of a nudge and a variety of uses for nudging (e.g., Schmidt 2017, for review look: Barton and Grüne-Yanoff 2015). For example, there are definitions of "nudge" that limits them to those interventions that target cognitively shallow processes and leave open the possibility for the person being nudged not doing what they are nudged to do (Saghai 2013). Nudges can be used to lead people to healthier lifestyles or be more careful while urinating, among m any other things.

The definition of a nudge operative in this article is close to Sunstein's above in that the interventions proposed here aim to stimulate conscious reflection without compromising their target's ability to choose to do otherwise. It is important to note, however, that while these nudges engage conscious high-level cognition, their ultimate aim is to influence the operation of lower-level mechanisms. For example, some proposed nudges prompt individuals to ask questions about their choices, values, and lives with the aim of improving their disposition to take the perspective of others.

It should also be noted that all of the proposed nudges could be implemented in a computer program that is not embodied in a robotic chassis. Given this, the present proposal can be counted among other computer-aided moral improvement technologies proposed in the literature. Among these is the artificial moral advisor proposed by Giubilini and Savulescu (Giubilini and Savulescu 2018) that is designed to promote a capacity for reflective equilibrium. There is also the ambient intelligence system proposed by Savulescu and Maslen (Savulescu and Maslen 2015) that can prompt its users about their emotional state, whenever they find themselves in a morally-charged situation, and even give advice based on previously selected values, such as concern for justice or honesty. There are also proposals that stress the importance of dialogue and argument, so an artificial moral reasoner (Klincewicz 2016; Lara and Deckers 2019). This raises the question of what putting nudges into a robot adds to those others computer-based solutions.

Embodiment in a chassis—ideally more-or-less humanoid—affords advantages that are not achievable with a phone app or computer application. As anyone that has interacted with a humanoid robot can attest, an appropriately designed

robot can engage us emotionally (Asada et al. 2009). A robot's smile, nod, or other programmed gesture can elicit a strong response. This offers an opportunity to target affective mechanisms that comprise moral cognition more directly. Other computer-based proposals for moral improvement that do not rely on a chassis do not afford this possibility. Besides, robotic nudges have already been shown to be effective in contexts with a moral dimension, e.g., when the robot resembles a person with authority (Aroyo et al. 2018) and in an ultimatum game paradigm (Di Dio et al. 2019).

The second major difference between the present proposal and other existing proposals for moral improvement with computing technology is that the nudges proposed here are inspired by Stoic practices. These practices have had a long track-record of being useful to politicians, generals, orators, and other non-philosophers in Ancient Greece and Rome. Practical advice that we inherited from this tradition has also been re-purposed in contemporary cognitive-behavioral therapy (CBT) (Clark and Egan 2015), with relative success. The link between Stoic practices and CBT is acknowledged by contemporary scholars as well as the founders of CBT themselves (for review look: Murguia and Diaz 2015; Robertson 2016; Beck 1979). One may even say that CBT continues with the Stoic idea that through appropriately directed conscious reflection we can help improve the overall quality of our psychological functioning.

Systematic reviews of meta-analyses of CBT studies consistently conclude that it is effective in treating a variety of psychological conditions, such as anxiety and depression (Beck 1991; Butler et al. 2006; Kazantzis et al. 2018). The roots of CBT practices in Stoic practices suggests then that we may be cautiously optimistic about the efficacy of Stoic practices themselves. If CBT is successful, then we have indirect evidence that Stoic practices may be effective for the purposes that they were originally intended for, namely, changing moral behavior. This reliance on Stoic practices to facilitate moral improvement is the second major difference between the present proposal and those present in the literature.

Section 2 of this article critically reviews some existing proposals of robotic moral nudgers. Section 3 reviews evidence from across the cognitive sciences to make the case for a capacity approach to nudging towards moral improvement. Section 4.1 presents a selection of Stoic practices that can inform engineering strategies that would help develop moral capacities. Section 4.2 develops the Stoic assumption that developing those capacities is a promising way towards moral improvement. Section 5 states and answers an objection to the development of robots that can morally improve through nudges, which is that there are less expensive

and more effective ways of stimulating moral development and moral improvement. In response, robotic nudgers offer some unique possibilities that are not available through other means, thus at least partially justifying the cost of their design and development.

2. Moral Improvement with Robotic Nudges

The semantics of 'improvement,' 'enhancement,' and 'therapy' are closely linked and there are several ways of differentiating between their meanings. For example, the distinction between 'therapy' and 'enhancement' is particularly important in context of medicine:

> In broad terms, therapy aims to fix something that has gone wrong, by curing specific diseases or injuries, while enhancement interventions aim to improve the state of an organism beyond its normal healthy state. (Bostrom and Roache 2008, 120)

On this view, therapy and enhancement can be thought of as falling under the more general concept of improvement. Therapy aims at improvement to normal, while enhancement aims at improvement beyond normal.

One problem with this way of making the distinction between different types of improvement is that it assumes some notion of normality, which inevitably imports statistical or normative assumptions about human nature. In a medical context this burden may sometimes be worth accepting. Prima facie, a medical professional should treat the infirm, but does not have an equal obligation to enhance the healthy. So, one may argue, the therapy/enhancement distinction may help to justly distribute limited medical resources in society (Kluge 2007; Daniels 1996).

In context of moral behavior, the burden of relying on the notion of a statistically normal person or some norm of an ideal person is more difficult to justify. It is not at all clear what it means to be morally normal and further whether normality should play any role in defining what counts as moral improvement (Daniels 2009; Kingma 2007). For one, it is not obvious why we should not always strive to go beyond what is morally normal for our species. The often-cited motivation for wholesale moral improvement, namely, the possibility of global catastrophe caused by our species' tribal moral psychology, seems to demand that we do in fact go beyond what is normal (Persson and Savulescu 2012, 12). Secondly, people have complex moral psychologies that can differ substantially across individuals. For example, one person may be above average in their concern for social justice, but find it hard to care about particular individuals, while another could be the opposite. Averaging

concern for social justice and care for individuals across these two persons would give us a distorted picture of what a normal moral psychology is like.

Given these complications, in the remainder of the article the more general notion of *moral improvement* will be used to refer to all the strategies that can inform moral nudges embodied in robots. This is meant to underline the distinction between the more general notion of moral improvement and *moral enhancement* as it is used in the literature to refer to improvement techniques that take advantage of biological, genetic, and neural interventions (Douglas 2008). Moral improvement can also be used to signify moral therapy, with its medical connotations, for special cases where moral improvement techniques can be useful is helping those whose moral development has in some way been derailed. Candidates for moral improvement as therapy may be, for example, young adults that display personality traits from the so-called dark triad (Paulhus and Williams 2002), child soldiers (Wessells 2006), or compulsive liars (Dike, Baranoski, and Griffith 2005). In similar vein, Stoic practices will be characterized as potential methods of moral improvement in this broad sense, thus ignoring their specific aims within Stoic philosophy, or characteristics of individuals and groups that may be targets for them.

Robots that could be used in the task of moral improvement have been proposed by Jason Borenstein and Ronald Arkin (Borenstein and Arkin 2016b; Borenstein and Arkin 2016a). Their idea is to use the robot as a source of nudges, which "could range from the sophisticated and subtle (such as crossing its arms and tilting its head) to the blunt and obvious (such as voicing the phrase 'please stop doing that')" (Borenstein and Arkin 2016b, 35).

Borenstein and Arkin argue that John Rawls's theory of justice as fairness (Rawls 2009; Rawls 2001) can serve as the theoretical background against which specific moral nudges could be programmed into the robot. It is worth quoting them at length to illustrate how this would work:

> [Robotic nudges] could nurture ''inequality aversion'' in young children by reinforcing proper social norms and etiquette during playtime. For example, a robotic companion could smile or display other social cues that encourage the sharing of toys between playmates. Along these lines, the robot could mimic expressions of disappointment if a child refuses to share. Furthermore, the robot could nudge a child to interact with other children with whom he/she is not as used to engaging in the effort to avoid ''parochialism.'' (Borenstein and Arkin 2016b, 40)

According to Borenstein and Arkin, parochialism and lack of concern for inequality are not conducive to concern for social justice on the conception of justice adopted by Rawls.

What motivates Borenstein and Arkin in making this connection is Rawls's second principle of justice: "social and economic inequalities, for example inequalities of wealth and authority, are just only if they result in compensating benefits for everyone, and in particular for the least advantaged members of society" (Rawls 2009, 13). They stipulate that encouraging inequality aversion and parochialism aversion with robotic nudges would amount to an indirect method of inculcating people with concern for Rawls's second principle. If this is right, improving those attitudes via nudges would amount to moral improvement, from a Rawlsian perspective.

One could disagree with this stipulation and point out that inequality aversion and aversion to parochialism are not exclusively Rawlsian attitudes and that many ethical frameworks are compatible with them.[2] While Borenstein and Arkin acknowledge that they fall short of showing why it is that these attitudes as opposed to other attitudes should be encouraged, they, arguably, do not sufficiently explain why they take them to be Rawlsian as opposed to Kantian, Aristotelian, or corresponding to another moral theory. The role of theoretical assumptions in moral improvement proposals, including moral bio-enhancement, is a more general problem that is unlikely to be solved soon (for review: Specker et al. 2014).

While Borenstein and Arkin's idea may be hostage to a philosophical debate, the practical viability of their proposal leaves room for exploration of alternatives. A particularly promising strategy is a more theory-neutral approach, which focuses on aspects of moral psychology, rather than on specific moral attitudes or beliefs. Beliefs can be understood as either occurrent or dispositional mental states that are distinguished from other mental states by featuring an assertoric attitude. On this view, to believe something is to hold a certain attitude towards a content, which is typically expressed as a proposition that can be either true or false. Beliefs provide reasons, including those relevant to moral behavior and moral decision making.

Beliefs can be distinguished from capacities, which are understood as the sum of the psychological and neural mechanisms that make it possible for agents to have mental states, such as beliefs. On this view, high-level psychological phenomena, such as believing that *p*, are understood to be decomposable through functional analysis into its component parts and realizers (Cummins 1975; Craver 2007). This is a gross oversimplification that rides roughshod over a great deal of discussion in the philosophy of mind on the nature of beliefs. However, in context

of this article it allows us to distinguish proposals that focus on beliefs and attitudes, such as Borenstein and Arkin's, from those that target the capacity to have such beliefs and its constituent mechanisms.

Proposals that fall into the latter category include those that aim to develop particular dispositions (Jebari 2014), reliability (Schaefer and Savulescu 2019), self-interest and cognitive capacity (Ahlskog 2017), or cognitive-affective mechanisms that underlie moral dispositions (Klincewicz, Frank, and Sokólska 2018). These approaches all share the assumption that beliefs or attitudes are not the appropriate targets for moral improvement, but the underlying capacities for having such beliefs are appropriate. One advantage of these views is that they do not have to wait for moral philosophy to deliver a verdict on which beliefs or attitudes are appropriate. Instead, they can focus on, for example, the psychological mechanism responsible for the ability to have moral attitudes in general. This mechanism may be important to moral behavior and moral decision-making regardless of which normative theory or metaethical position turns out to be right.

Focusing on mechanisms and capacities is not in itself completely theory-neutral. Each time a mechanism is considered as important to moral decision-making or moral behavior, assumptions about moral psychology and moral behavior are smuggled in. These assumptions will aim to capture the surface phenomenon of human moral behavior by answering questions such as: is this a morally-charged decision? Does this action have a morally significant consequence? Or, what distinguishes moral behavior from non-moral behavior? Arguably, answers to these sorts of questions will be roughly the same across first-order normative theories and metaethical positions (for example, see: Sterba 2004).

Borenstein and Arkin consider a capacities approach in a follow-up paper, which describes robotic nudges that encourage "performance of charitable acts" and "promote the good of society" via nurturing empathy (Borenstein and Arkin 2016a). The upshot of their discussion is that

> [e]quipped with the knowledge of affective computing and other disciplines, roboticists could exert profound power over the humans that interact with robots. Thus, we sought to explore a possible design pathway whereby robots would seek to nurture a user's empathy toward other human beings and more specifically, nudge the user toward the performance of charitable acts. (Borenstein and Arkin 2016a, 8)

This sets a modest and achievable goal of improving morally relevant capacities with existing solutions in affective computing (Picard 1997; Calvo et al. 2014).

For example, we can imagine a pattern recognizer that can detect when people take on the perspective of others or become angry and pass this information to a robot, which can use that information to generate an appropriate nudge.

The problem with that proposal is that it is not clear why performance of charitable acts and promoting the good of society should be encouraged by the robotic nudger as opposed to something else. Secondly, empathy, which Borenstein and Arkin focus on, has several competing theoretical accounts in psychology and philosophy (for review: Cuff et al. 2016; Stueber 2017). Furthermore, empathy's role in moral cognition has advocates (Hoffman 2001), but also detractors (Maibom 2014; Isserow 2015; Huebner 2015). There are also reasons to think that empathy is, at times, not morally desirable at all. Paul Bloom, among others, points out that empathy introduces biases into moral deliberation that cause us to put too much value on those people we easily empathize with and too little on those with which we do not (Bloom 2017).

The controversial nature of empathy puts pressure on Borenstein and Arkin's version of the capacities approach to moral improvement via robotic nudges. It also suggests a posture of caution about any particular moral capacity or set of mechanisms being the key to moral improvement across all individuals. There may be significant individual and group differences that demand a more nuanced approach.

Even with these criticisms in mind, the robots Borenstein and Arkin describe can serve an important role in stimulating the development of psychological capacities that underlie moral decision-making, and in particular those capacities that are connected to emotions. For example, helping affective perspective taking develop in young people, especially those in risky environments, such as conflict zones or juvenile detention centers, may be a role that robotic nudgers can fulfill uniquely. In sum, Borenstein and Arkin's arguments make it possible to think about social robots as a source of nudges towards moral improvement, the strategies they adopt in that task, that is, the Rawlsian framework and a moral capacities approach that focuses on empathy, are not optimal. A good source of possible alternative capacities to focus on is the wealth of research on moral development across neuroscience and developmental psychology.

3. Moral Development

The idea that moral development is connected to psychological development is relatively new. It was made scientifically prominent by such luminaries of psychology as Emile Durkheim, Pierre Piaget and eventually Lawrence Kohlberg (Snarey and Samuelson 2008). On Kohlberg's view, human beings move through six stages of

moral maturity, which start from a child's conception of the relationship between obedience and punishment all the way to abstract reasoning that involves duties and imperatives (Kohlberg 1984). If that were right, moral improvement would be a matter of moving through these stages as quickly as possible until one reaches the last, which eminently involves cognitive capacities, such as abstract reasoning.

Kohlberg's view has come under significant criticism largely on the basis of individual and group differences in moral reasoning (Hwang 2015; Snarey 1985). There is a significant amount of evidence that people that Kohlberg's view would place at the same level of moral development may use significantly different strategies and psychological means to make moral decisions (Krebs and Denton 2005). This suggests that people can differ significantly with respect to their psychological development yet be on the same stage of moral development or vice versa.

Kohlberg's emphasis on justice, rights, and reasoning has also come under significant criticism from Carol Gilligan (1977). Gilligan points out that a capacity for care is an equally important component of moral cognition and development that Kohlberg's theory neglects. Gilligan's approach has mounting support in biology, neuroscience, and social sciences, which all support the view that affective mechanisms neglected by Kohlberg's theory play an important role in moral cognition. Current research into moral development encompasses this more nuanced approach (Lapsley and Narvaez 2005; Lapsley and Carlo 2014; Walker 2004), which characterizes moral development as an aspect of a multifaceted process of psychological development (Johnson 2015). Emotional development plays an important role in a child's development of moral capacities (Pizarro, Detweiler-Bedell, and Bloom 2006). On this view, the level of an individual's moral development is typically assessed by measures of pro-social attitudes, ability to engage in moral reasoning, and other, related capacities, such as capacity for care.

What many theorists seem to agree on is that moral cognition is realized by a complex network of cognitive and affective mechanisms, rather than by a single, domain-specific mechanism (Greene 2015) and that moral development is dynamic (Van Bavel, FeldmanHall, and Mende-Siedlecki 2015). This more complex picture according to which moral development involves distinct stages connected to the maturation of distinct emotional and cognitive capacities (Cowell and Decety 2015) diverges significantly from the Kohlbergian conception. It is also supported by empirical studies, which demonstrate that distinct neural mechanisms are involved in moral behavior at different ages. Affective mechanisms are shown to underlie moral dispositions in infants (Vaish, Grossmann, and Woodward 2008; Hamlin, Wynn, and Bloom 2010) and likely involve subcortical structures, but not

to the same degree in young people and adults (Hyde, Shaw, and Hariri 2013). A similar conclusion is supported by lesion studies of younger people with damage to the ventromedial prefrontal cortex and amygdala, which display significant problems in processing morally laden stimuli and social emotions (Gupta, Tranel, and Duff 2012). Similar damage later in life affects these processes significantly less. Studies of pathologies, especially in psychopaths, corroborate these conclusions (Kennett and Fine 2008).

Decety and Howard summarize the way in which the amygdala/superior temporal sulcus/ventromedial prefrontal cortex neural network is involved in moral cognition (Decety and Howard 2013, 52–53; Decety and Cowell 2014). These structures are known to be responsible for integrating information from a variety of other brain regions, including those responsible for affect and motivation (Motzkin et al. 2015; Jalbrzikowski et al. 2017). With age the role of each component in the network changes in importance, with emphasis often moving from affective processing to cognitive processing.

In general, empirical studies on moral development strongly suggest that the moral psychology of children and young adults is significantly different from that of adults and that these differences are both anatomical and functional. This view is further supported by what we know about the way that people develop other aspects of their social cognition. For example, young adults do not always have a fully developed prefrontal cortex, critical to decision-making in general, and moral decision-making in particular. Young adults nonetheless can and do make moral decisions, which further supports the view that their moral psychologies are typically realized by a different complex of psychological capacities than that of adults (Caballero, Granberg, and Tseng 2016). Moral improvement interventions should be sensitive to the fact that affective mechanisms, specifically those involved in empathy, are important at a young age. Higher-order cognition seems to become more important later in development.

The design of moral improvement interventions is further complicated by the fact that a child's moral development can be derailed by many things, which can be obscure to a professional. Inadequate pro-social stimulation (Padilla-Walker 2014), brain lesions (Taber-Thomas et al. 2014), trauma (Narvaez and Lapsley 2014), and a range of other factors can all impact moral development (Gibbs 2010). What further compounds the difficulty is that we have strong evidence that changes in behavior are accompanied by neuroanatomical changes (Sandi and Haller 2015; Glenn and Raine 2014).[3]

These observations about moral development have significant implications for strategies of moral improvement. Most importantly, moral improvement interventions should not take a one-size-fits-all approach that targets one specific capacity or neural realizer of moral cognition. Ideally, they would take all the relevant mechanisms into account or focus on the unique ways in which moral cognition is realized in each individual. Moral therapy, if there were such a clinical practice, would then be an interdisciplinary domain mainly for clinical and developmental psychologists.

It is here that focusing on capacities and constituent mechanisms of moral cognition, as opposed to specific beliefs and attitudes, proves its superiority as an approach to moral improvement. Take, for example, the psychological mechanism responsible for affective perspective taking, which underlies some of the moral decisions that people may make about others. Focusing on improvement of the functioning of that particular mechanism will likely lead to the development of an overall improvement in concern for others, including those that are not like us. Furthermore, this approach offers a way generate empirically testable hypotheses, such as the one about affective perspective taking and care, that can be tested in particular individuals and in groups.

Robots designed to deliver interventions that target specific mechanisms relevant to moral development may be a way to supplement the work of therapists and give them a powerful tool. The robotic nudgers proposed by Borenstein and Arkin (Borenstein and Arkin 2016a) could, as they speculate, facilitate emotional development relevant to moral cognition in children and possibly also in adults. But instead of being designed to promote Rawlsian principles or empathy they would need to be designed with strategies for improvement of other psychological capacities relevant to moral behavior and moral decision-making. Concrete strategies to this end can be found in Stoic moral theory, which has recently been suggested as an alternative to the typical utilitarian/deontological strategies considered for moral AI systems (Murray 2017).

4. Stoicism and Moral Improvement

4.1. Stoic Practice

Stoicism was one of the major schools of ancient philosophy with a developed metaphysics, epistemology, philosophy of language, logic, and ethics (Schofield 2003). Not much of this remains, except for fragments, which makes a thorough reconstruction of Stoicism difficult. What is best preserved is practical advice,

presumably grounded in all that theory, and passed on through the writings of statesmen and orators, such as Marcus Aurelius and Cicero.[4] The selection of these practices below is not exhaustive, but for each a brief speculative sketch about how it can be embodied in a social robot is provided.

1. *Assessment of control*. Stoics recommend that people learn to distinguish between things that they have full control over from those that they have partial or no control over. Attempting to influence things that we have no or little control over is usually futile. The idea here is that once we learn to discern them, we will commit less time and mental energy to things that can only frustrate us (Epictetus 1983, 1).

A robotic nudger can be designed to help its user develop the ability to discriminate what is up to them and what is not. For example, the robot could nudge its user to reflect on the amount of control that they have over what they hear on the news. Or the nudger could make the relevant distinctions themselves and then confront them with the distinctions that its user makes.

To work well, the robotic nudger would have to be designed with the ability to discern appropriate times and topics for this assessment. It is not clear what criteria should be used in this task, but a good first pass may be a recorded history of topics, situations, and bodily responses of its user. When confronted with something that is likely to elicit negative emotions, the robotic companion could preemptively prompt to reflect on the amount of control one has over that something.

2. *Imagining calamity* involves visualizing or thinking about the loss of something dear. Epictetus recommends:

> If, for example, you are fond of a specific ceramic cup, remind yourself that it is only ceramic cups in general of which you are fond. Then, if it breaks, you will not be disturbed. (Epictetus 1983, 3)

The goal of this practice is to maintain one's appreciation for having the ceramic cup and also combat dispositions to cling to it. A robotic companion could prompt its user to engage in imagining the loss of something dear by asking relevant questions, such as "what if your ceramic cup was broken?"

Taken to extremes this practice can be jarring. Epictetus suggests that "if you kiss your child, or your wife, say that you only kiss things which are human, and thus you will not be disturbed if either of them dies" (Epictetus 1983, 3). It is difficult to expect anything good to come from a robot prompting a happily married father with the question "what if your wife and children were dead?" To avoid such extremes, the robot should be able make the relevant distinctions among

things that are appropriate for this technique and those that are out of bounds. To accomplish this, the engineer may want to create a mechanism for creating a hierarchy of values and things that are dear to the user. This hierarchy may then be categorized by a psychologist, perhaps with an eye to principles of cognitive-behavioral therapy. This would go far is avoiding a situation in which the robot would become a nuisance rather than a useful nudger.

3. *Preparation for and review of the day*. This practice has several different across each prominent Stoic thinker. Here is an excerpt from Marcus Aurelius's advice for the morning:

> Begin the morning by saying to thyself, I shall meet with the busy-body, the ungrateful, arrogant, deceitful, envious, unsocial. All these things happen to them by reason of their ignorance of what is good and evil. (Aurelius 2013, 2.1)

And Seneca's advice for going to sleep:

> I shall keep watching myself continually, and—a most useful habit—shall review each day. For this is what makes us wicked: that no one of us looks back over his own life. Our thoughts are devoted only to what we are about to do. And yet our plans for the future always depend on the past. (Seneca 2001, 83.2)

The goal of these recommendations is the same: to be reminded of one's place in the world and to foster a mindful and tranquil disposition in the face of frustration and ill-will of others.

There is no one-size-fits-all method of preparing for and reviewing the day. There are also probably many ways in which a robotic companion could nudge its user to engage in preparation for and review of the day. The simplest would be to prompt with "have you reviewed your day?" But we could also imagine a more involved method with leading questions or a template, which would make it easier for the user to reflect on relevant aspects of the day, such as lunch with a colleague, small frustrations with a partner, etc.

4. *Staying in the present*. Reviewing one's day does not mean the same as ruminating about what could have been, which the Stoics consider to be dangerous. To counteract the tendency to ruminate, the Stoics recommend staying busy. For example, Seneca tells us:

> The present is short, the future is doubtful, the past is certain. For this last is the one over which Fortune has lost her power, which cannot be brought

> back to anyone's control. But this is what preoccupied people lose: for they have no time to look back at their past, and even if they did, it is not pleasant to recall activities they are ashamed of. (Seneca 2004, 2.1)

Seneca's observation recommends we occupy ourselves, so we would spend less time thinking about what could have been. Confirming this recommendation, recent experiments suggest that indeed ruminating on negative aspects of the past correlates with lowered psychological well-being (Blouin-Hudon and Zelenski 2016).

A general recommendation, such as "stay in the present" or "stay busy" does not straightforwardly lead to specific practices. But even a general recommendation, such as to be mindful, can be suggestive to a range of therapeutic strategies (Germer, Siegel, and Fulton 2016). Cognitive therapists use cognates of "stay in the present" to remind their clients to not ruminate, for example. Being mindful can also refer to directed attention, such as in some forms of meditation. Prompts to be mindful can be implemented in a robotic nudger to suggest any one of these, including meditation or directed attention.

Natural language analysis from artificial intelligence (Hirschberg and Manning 2015) could be especially useful here. We could imagine a robotic nudger that can identify negative ruminating on the past in its user's language use (Wildschut et al. 2006). The input for this analysis could be the linguistic corpus produced in strategies (1), (2), and (3), but it could also be interviews or a therapist report. The output would of course be an appropriate nudge or dialogue that suggests a positive aspect of the situation one tends to ruminate about.

5. *Controlling one's emotions*. Negative emotions attract a lot of the Stoics' attention. Cicero compares negative emotions to insanity (Cicero 2002, 3.23, 3.5) and recommends philosophy—understood here to be engaging one's reasoning ability—as cure. Similarly, Marcus Aurelius offers a wide range of examples of how to think about the emotions and their causes (Aurelius 2013, 9, 11). Most Stoics recommend thinking through the unreasonableness of one's emotional perturbations and expect this practice to eventually change one's dispositions to have them. With the dispositions changed, the Stoic will presumably stop getting angry at, say, their noisy neighbor or at things in general.

There is a lot that a robotic nudger could do to encourage reflection on the causes of one's emotional perturbances. Most of the strategies discussed above are, arguably, versions of this more general strategy. This is also where Borenstein and Arkin's own proposal may not need much else than they already provide.

What may make it even better is equipping the robotic nudger with the ability to discern emotions in the user and perhaps even anticipate them, so as to be able to interrupt them or facilitate thinking through them. Insights and work in affective computing could be paramount to achieving this end.

To sum up, this article has so far outlined the theoretical framework for designing robotic nudges for moral improvement based on Stoic practices and on insights from psychology and developmental science. Stoic practices were originally developed with moral improvement in mind, but have proved successful in cognitive-behavioral therapy. This success indirectly supports some optimism about the success of Stoic practices themselves. However, it does not allay any worry that may arise about the potential success of any individual Stoic practice embodied in a robotic nudge. The practical and ethical issues that bear on each practice as embodied in a robotic nudger is beyond the scope of this article and should instead be address by empirical work at the intersection of psychology, human-robot interaction studies, and moral philosophy. We cannot speculate *a priori* about the potential success of any of these at the present moment, but we can be hopeful, given the success of Stoic practices channeled via cognitive-behavioral therapy.

4.2. Stoic Doctrine of oikeiosis

Most ancient moral theories, including the Stoic one, conceives of being moral as a matter of a person reaching a certain state. For the Stoics in particular this state is *oikeiosis*—an enlargement of what one takes to be his or her own achieved by maximally expanding the range of one's concern from oneself to others by appropriate exercise of reason (Pembroke 1971). Appropriate exercise of reason can mean many things and practices like the ones listed in Section 4.1 are all examples of it. This idea can also be understood as a shorthand for a state of moral virtue marked by the four traditional ancient virtues: temperance, courage, wisdom, and justice. The key to reaching oikeiosis are the emotions, and specifically, an attitude towards emotions that Stoic practices aim to cultivate.

There are several intermediate stages before maximal oikeiosis is attained and in all but the first stage the key to achieving success is making rational choices based on sound logic. Cicero sketches out five such stages and marks out the role of reason in each one:

> The first 'proper function' . . . is to preserve oneself in one's natural constitution; the second is to seize hold of the things that accord with nature and to banish their opposites. Once this procedure of selection and rejection has been discovered, the next consequence is selection exercised with

> proper functioning; then such selection is performed continuously; finally, selection which is absolutely consistent and in full accordance with nature. (Cicero 2001, 3.20–1)

On this view, a person's actual choice and natural choice become ever more in sync until they become the same when one reaches maximal oikeiosis. At that point, a person is in harmony with nature and most free.

The goal of being in harmony with nature cannot be fully appreciated without a better understanding of Stoic metaphysics. As Gisela Striker explains:

> The Stoic conception of the end does not arise as a natural continuation of one's concern for self-preservation, but rather as the result of one's reflection upon the way nature has arranged human behavior in the context of an admirable cosmic order. (Striker 1991, 230)

According to the Stoics, this admirable cosmic order is deterministic and beyond what a human mind can fully grasp. Nonetheless, even though the whole mechanism is incomprehensible, one's place in it can be made apparent by practicing Stoic ethics and maximizing oikeiosis. The idea of an admirable cosmic order is unlikely to motivate many people to engage in the project of maximizing oikeiosis or to become a Stoic. This is perhaps why the final stage in Cicero's description of "selection which is absolutely consistent and in full accordance with nature" is also considered by the Stoics to be very difficult to achieve. It is perhaps also why Stoic ethics does not have many adherents today.

Those unmoved by the idea of being a gear in the mechanism of the admirable cosmic order may be more moved by the characterization of oikeiosis in psychological terms.[5] Hierocles provides this description as the final stage of a successive development towards moral virtue. Stages in this development are marked by extending the range of concern from oneself, to parents, siblings, wife, and children, then to uncles, aunts, grandparents, nephews, nieces and cousins, to other relatives, to local residents, fellow tribesmen, etc. "The outermost and largest circle, which encompasses all the rest, is that of the whole human race" (Long and Sedley 1987, citing Stobaeus 4.671.7–4.673.11).

On Hierocles's account, the key to maximizing oikeiosis is reaching a mental state in which one's concern extends to an ever-wider range of people. This squares well with Malcolm Schofield's suggestion that "*oikeiosis* theory proposes that concern for others depends on *identifying* with them in some way or to some extent. . . . It involves the disposition to adopt the other's point of view" (Schofield

1995, 196; Schofield's emphasis). Identification with others and adopting their point of view is what we oftentimes call empathizing. However, it is important to note that the contemporary psychological category that best fits the Stoics' conception is not an emotive attitude, nor a belief, but something akin to perception.

To avoid anachronisms, the relevant mental state can be cautiously analyzed as having two parts: a mental attitude and (propositional) content. We can thus characterize Hierocles's idea of the expanding oikeiosis with a shorthand 'O' attitude:

A) Sam Os 'Sam'

B) Sam Os 'Sam's mother'

C) Sam Os 'All the members of Sam's species'

The O-attitude is distinguished from other propositional attitudes, such as believing or hoping, by the causal role it plays in the mental life of the person that has it—it is inherently motivational. Its essential feature is to bring about concern for the well-being of whatever it is directed at by identifying with them, but it is not identical with that concern.

On this model, as the content of one's O-attitudes becomes more general, as is the case from (A) to (B) to (C), one progresses towards maximal oikeiosis. Hierocles's idea is that we should end up with a stable, emotion-like attitude, rather than a mere belief with ever more general content. But reason and logic are still essential, as they are the basis for all the Stoic exercises mentioned in Section 4.1. Just as with Cicero's stages in understanding the cosmic order, Hierocles thinks that a Stoic uses reason and logic to move towards each successive stage.

If the Stoics are right about any of this, an individual's moral progress can be measured independently from the development of ability to engage in any of the practices and exercises they suggest. A measure of this development is the scope of concern with others, that is, the level of generality of the O-attitude. The more cosmopolitan we are, the more virtuous we are. The more parochial we are, the less virtuous we are. Importantly, this idea applies to everyone, not only to people that have had their moral development derailed, such as child soldiers or compulsive liars. The Stoics emphasized that their practical advice and exercises can be used by people of all ages, regardless of their education, ability or socioeconomic status. What this suggests, is that if the robotic nudger proposed here can help people with problems in their moral development, they could equally help with people without such problems. This gives genuine hope for moral improvement that relies on robotic nudging technology.

5. Should We Bother Building Robotic Nudgers for Moral Improvement?

Long-term use of social robots raises ethical concerns that should be addressed in any future development (Coeckelbergh et al. 2016). Programming ethics or morality into robots is also subject to a number of "rookie mistakes" that are the consequence of the relative lack of clarity about what ethics and morality are among engineers (Gordon 2019). While the promise of long-term use of robots in care, therapy, and companionship is better than ever, it also faces significant limitations (Cabibihan et al. 2013; Leite, Martinho, and Paiva 2013; Frank 2019). The main limitation for their widespread use is practical: it remains prohibitively costly to design and develop humanoid robots that work as intended. This last concern is particularly pressing for robotic nudgers, since it is not clear what the benefit of using them is, as compared to other solutions.

The development, deployment, and use of sophisticated robotic nudgers would indeed involve a great amount of research and resources. Furthermore, it remains unclear whether they can ever achieve the kind of sophistication that would be necessary to aid the developmentally challenged or to facilitate moral improvement in children, or in adults. Finally, the possibility of implementing Stoic practices and their potential effectiveness in therapy is speculative and support for it indirect, since it depends on success of cognitive-behavioral techniques. On the other hand, it seems just as likely that the same amount of research and resources put towards caretakers, teachers, and therapists would achieve as much, if not more, to allay derailed moral development or to morally improve. In short, humans—Stoic or otherwise—may do as well as the robots and for less, which means that we lack a substantive reason to develop this technology.

In reply, one should concede that robotic nudgers cannot replace caretakers, teachers, and therapists, at the present time. They can nonetheless be an effective supplement to what all of these professionals can accomplish on their own. There are also independent reasons to think that this technology could provide unique new opportunities that could not be otherwise realized.

Robots can help people with developmental problems develop an emotional bond and to engage emotionally in ways not achievable with disembodied computer programs or even human caretakers. The affordance of emotional engagement is perhaps also why robots have been used in behavioral therapy with relative success and why robots designed to help the elderly or people with cognitive impairments have been the subject of significant recent research and development (Robinson, MacDonald, and Broadbent 2014; Doering et al. 2016; Gross et al. 2016; Gross et

al. 2015; Fischinger et al. 2016; Magnenat-Thalmann and Zhang 2014). Existing solutions achieve a modest amount of success in the tasks for which they were designed (Peri et al. 2016; Ahn et al. 2014). For example, a recent empirical study shows that lonely seniors are more likely to take a walk when accompanied by a robotic companion (Karunarathne et al. 2018). Robots can also significantly help in second-language acquisition (Belpaeme et al. 2015).

The effectiveness of robots in therapy and education of clinical populations has been demonstrated in several empirical studies. First, the robot Kaspar facilitated learning cause-and-effect relationships and awareness of self in children with Autism Spectrum Disorder (ASD) through tactile play scenarios (Robins and Dautenhahn 2014). Models of kinematics for therapy robots for children with autism engage them with great success (Ge, Park, and Howard 2016). In other words, children with autism and possibly also other developmental problems (Standen et al. 2016), are likely to engage robots in play and disengage from their often closed-in inner lives. They can do this when they would not with other people.

Second, it been shown that children with ASD are generally more likely to engage robots in joint attention (JA):

> [C]hildren with ASD who interacted with the robot had better outcomes in terms of JA than the children who interacted with a human agent during all sessions and exhibited improved performance in a JA task with human after interacting with the robot. (Kumazaki et al. 2018, 6)

In context of ASD robots are unique, in that they can engage people that are otherwise disengaged from human caregivers or instructors. Robots have been shown to be uniquely useful in helping people with developmental problems, such as autism (Costa et al. 2015) and even Down syndrome (Lehmann et al. 2014) where human interaction has failed.

The unique opportunity that robotic nudgers give that caretakers do not is their being robots and not people. They remain emotionally engaging, because of their ability to elicit emotional responses. These unique features of therapy robots may equally help interventions on people affected by problems in moral development and those seeking moral improvement. A robotic nudge that embodies Stoic practices would be more likely to work for such people than a similar prompt from a human: people disposed to violence, aggression, or anti-social behaviors, are so disposed towards people, not robots. This means that they may be less susceptible to actualize negative dispositions when dealing with situations analogous or even identical to those they could find with people. This in turn could lead to a change

in their dispositions in ways that would otherwise be impossible. This is a unique added value afforded only by non-human animals, which cannot engage in the sort of human-like behavior that robots can.

Another unique feature of social robots is that the amount of time that a potential user spends with them can far outrun what is possible with another person. Long-term care of developmentally disabled can consume a large amount of resources, which may not always be available. When we consider these costs, we should realize that a similar amount of resources may never be diverted to those that may need moral rehabilitation or help in overcoming problems in their moral development. What is more likely is that help with their problems would be relegated to their social support networks, if they have any, or to the judicial system. Robots, such as the robotic nudgers proposed here, can provide long-term care for these populations. It remains an open question whether this sort of intervention can be achievable, the costs of building these sophisticated robotic nudgers notwithstanding.

6. Conclusion

Robotic nudgers can facilitate moral development or overcome problems with moral development, if their design focuses on the relevant psychological capacities. Stoic exercises, which have found their use in cognitive-behavioral therapy, are a good source of engineering strategies to implement as nudges in these robots. A further advantage of this approach is the possibility that it will also be sufficient for moral improvement through expansion of concern for others. The objection that robotic nudgers are too costly or unnecessary is blunted by considerations of the range of unique opportunities that robots provide, but it remains an open question whether these considerations are sufficient to justify their cost.

Notes

I would like to thank anonymous reviewers and special editors for the many helpful and constructive comments that made this article reach its potential. An earlier version of this article was presented at the Philosophy of Risk seminar organized by Sven Nyholm as a part of The Dutch Research School of Philosophy (OZSW) in the Technical University of Eindhoven where I received helpful feedback. Parts of this article were submitted as a seminar paper in Richard Sorabji's course on Stoic philosophy at the CUNY Graduate Center many years ago. Subsequent work on the paper was partially financed by the Polish National Science Centre (NCN) SONATA 9 Grant, PSP: K/PBD/000139 under decision UMO-2015/17/D/HS1/01705.

1. It is important to note that Stoicism should be understood within the historical context within which it was developed and should not be applied straightforwardly, without due changes, to the contemporary context. In this article Stoicism is presented without a sufficient discussion of its historical and dialectical context, but as a source of strategies that may, but need not be useful to contemporary moral concerns. Furthermore, some may worry that in order for these strategies to have any validity one should endorse Stoicism or that Stoicism has to be the right first-order normative theory. It should be stressed that this is not so. Stoic practical advice is useful in many contexts, even to those that are not Stoics.

2. I am grateful to an anonymous reviewer of this journal for drawing my attention to this point.

3. However, moral improvement and moral development is often treated as a domain of parents or other primary caregivers or religious authority. As a consequence, the people that may benefit most from professional help in overcoming problems in their moral development are sometimes the ones least likely to get it.

4. I found Section 4 of Irvine 2008 particularly helpful in making up this list of practices.

5. Along with Malcolm Schofield I suppose that "we can explain the two methodological approaches they reflect as complementary parts of a single coherent argumentative strategy" (Schofield 1995, 210), which has as its aim being in harmony with nature. The psychological and the metaphysical description together compose a comprehensive Stoic theory of compatibilism about free will, which is the view that the world is deterministic, but we can nonetheless be free to decide.

References

Ahlskog, Rafael. 2017. "Moral Enhancement Should Target Self-Interest and Cognitive Capacity." *Neuroethics* 10(3): 363–73. https://doi.org/10.1007/s12152-017-9331-x

Ahn, Ho Seok, Elizabeth Broadbent, Han Kuo, Chandan Datta, Rebecca Stafford, Hayley Robinson, Ngaire Kerse, Kathy Peri, and Bruce A. MacDonald. 2014. "Long-Term Study of a Healthcare Robot System for Senior People in Rest Homes, Hospitals, and a Dementia Unit." In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems Workshop on Assistive Robotics for Individuals with Disabilities: HRI Issues and Beyond*.

Aroyo, Alexander Mois, T. Kyohei, Tora Koyama, Hiroshi Takahashi, Francesco Rea, Alessandra Sciutti, Yuichiro Yoshikawa, Hiroshi Ishiguro, and Giulio Sandini. 2018. "Will People Morally Crack Under the Authority of a Famous Wicked Robot?" In *2018 27th IEEE International Symposium on Robot and Human In-*

teractive Communication (RO-MAN). The Institute of Electrical and Electronics Engineers, 27–31 August 2018: 35–42. https://doi.org/10.1109/ROMAN.2018.8525744

Asada, Minoru, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. 2009. "Cognitive Developmental Robotics: A Survey." *IEEE Transactions on Autonomous Mental Development* 1(1). The Institute of Electrical and Electronics Engineers, 28 April 2009: 12–34. https://doi.org/10.1109/TAMD.2009.2021702

Aurelius, Marcus. 2013. *Marcus Aurelius: Meditations*. Trans. Christopher Gill. Oxford: Oxford University Press.

Barton, Adrien, and Till Grüne-Yanoff. 2015. "From libertarian paternalism to nudging—and beyond." *Review of Philosophy and Psychology* 6(3): 341–59. https://doi.org/10.1007/s13164-015-0268-x

Beck, Aaron T. 1979. *Cognitive Therapy of Depression*. New York: Guilford Press.

Beck, Aaron T. 1991. "Cognitive Therapy: A 30-year Retrospective." *American Psychologist* 46(4): 368–75. https://doi.org/10.1037/0003-066X.46.4.368

Belpaeme, Tony, James Kennedy, Paul Baxter, Paul Vogt, Emiel E. J. Krahmer, Stefan Kopp, Kirsten Bergmann, Paul Leseman, Aylin C. Küntay, Tilbe Göksun, Amit K. Pandey, Rodolphe Gelin, Petra Koudelkova, and Tommy Deblieck. 2015. "L2TOR-Second Language Tutoring Using Social Robots." In *Proceedings of the ICSR 2015 WONDER Workshop*.

Bloom, Paul. 2017. *Against Empathy: The Case for Rational Compassion*. New York: Random House.

Blouin-Hudon, Eve-Marie C., and John M. Zelenski. 2016. "The Daydreamer: Exploring the Personality Underpinnings of Daydreaming Styles and Their Implications for Well-Being." *Consciousness and Cognition* 44: 114–29. https://doi.org/10.1016/j.concog.2016.07.007

Borenstein, Jason, and Ronald C Arkin. 2016a. "Nudging for Good: Robots and the Ethical Appropriateness of Nurturing Empathy and Charitable Behavior." *AI & Society* 32(4): 1–9. https://doi.org/10.1007/s00146-016-0684-1

Borenstein, Jason, and Ron Arkin. 2016b. "Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being." *Science and Engineering Ethics* 22(1): 31–46. https://doi.org/10.1007/s11948-015-9636-2

Bostrom, Nick, and Rebecca Roache. 2008. "Ethical Issues in Human Enhancement." In *New Waves in Applied Ethics*, ed. Jesper Ryberg, Thomas Petersen, and Clark Wolf, 120–52. London: Palgrave Macmillan.

Butler, Andrew C., Jason E. Chapman, Evan M. Forman, and Aaron T. Beck. 2006. "The Empirical Status of Cognitive-Behavioral Therapy: A Review of Meta-Analyses." *Clinical Psychology Review* 26(1): 17–31. https://doi.org/10.1016/j.cpr.2005.07.003

Caballero, Adriana, Rachel Granberg, and Kuei Y. Tseng. 2016. "Mechanisms Contributing to Prefrontal Cortex Maturation during Adolescence." *Neuroscience and Biobehavioral Reviews* 70: 4–12. https://doi.org/10.1016/j.neubiorev.2016.05.013

Cabibihan, John-John, Hifza Javed, Marcelo Ang, and Sharifah Mariam Aljunied. 2013. "Why Robots? A Survey on the Roles and Benefits of Social Robots in the Therapy of Children with Autism." *International Journal of Social Robotics* 5(4): 593–618. https://doi.org/10.1007/s12369-013-0202-2

Calvo, Rafael A., Sidney D'Mello, Jonathan Gratch, and Arvid Kappas. 2014. *The Oxford Handbook of Affective Computing*. New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199942237.013.040

Cicero, Marcus Tullius. 2001. *Cicero: On Moral Ends*. Trans. Julia Annas. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511803659

Cicero, Marcus Tullius. 2002. *Cicero on the Emotions: Tusculan Disputations 3 and 4*. Trans. Margaret Graver. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226305196.001.0001

Clark, Gavin I., and Sarah J. Egan. 2015. "The Socratic Method in Cognitive Behavioural Therapy: A Narrative Review." *Cognitive Therapy and Research* 39(6): 863–79. https://doi.org/10.1007/s10608-015-9707-3

Coeckelbergh, Mark, Cristina Pop, Ramona Simut, Andreea Peca, Sebastian Pintea, Daniel David, and Bram Vanderborght. 2016. "A Survey of Expectations about the Role of Robots in Robot-Assisted Therapy for Children with ASD: Ethical Acceptability, Trust, Sociability, Appearance, and Attachment." *Science and Engineering Ethics* 22(1): 47–65. https://doi.org/10.1007/s11948-015-9649-x

Costa, Sandra, Hagen Lehmann, Kerstin Dautenhahn, Ben Robins, and Filomena Soares. 2015. "Using a Humanoid Robot to Elicit Body Awareness and Appropriate Physical Interaction in Children with Autism." *International Journal of Social Robotics* 7(2): 265–78. https://doi:10.1007/s12369-014-0250-2

Cowell, Jason M., and Jean Decety. 2015. "Precursors to Morality in Development as a Complex Interplay between Neural, Socioenvironmental, and Behavioral Facets." *Proceedings of the National Academy of Sciences* 112(41): 12657–62. https://doi.org/10.1073/pnas.1508832112

Craver, Carl F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press. https://doi.org/10.1093/acprof:oso/9780199299317.003.0007

Cuff, Benjamin M. P., Sarah J. Brown, Laura Taylor, and Douglas J. Howat. 2016. "Empathy: A Review of the Concept." *Emotion Review* 8(2): 144–53. https://doi.org/10.1177/1754073914558466

Cummins, Robert. 1975. "Functional Analysis." *Journal of Philosophy* 72: 741–64. https://doi.org/10.2307/2024640

Daniels, Norman. 1996. "Justice, Fair Procedures, and the Goals of Medicine." *Hastings Center Report* 26(6): 10–12. https://doi.org/10.2307/3528745

Daniels, Norman. 2009. "Can Anyone Really Be Talking about Ethically Modifying Human Nature." In *Human Enhancement*, ed. Julian Savulescu and Nick Bostrom, 25–42. Oxford: Oxford University Press.

Decety, Jean, and Jason M. Cowell. 2014. "The Complex Relation between Morality and Empathy." *Trends in Cognitive Sciences* 18(7): 337–39. https://doi.org/10.1016/j.tics.2014.04.008

Decety, Jean, and Lauren H. Howard. 2013. "The Role of Affect in the Neurodevelopment of Morality." *Child Development Perspectives* 7(1): 49–54. https://doi.org/10.1111/cdep.12020

Di Dio, Cinzia, Federico Manzi, S. Itakura, Takayuki Kanda, Hiroshi Ishiguro, Davide Massaro, and Antonella Marchetti. 2019. "It Does Not Matter Who You Are: Fairness in Pre-schoolers Interacting with Human and Robotic Partners." *International Journal of Social Robotics*, online first. https://doi.org/10.1007/s12369-019-00528-9

Dike, Charles C., Madelon Baranoski, and Ezra E. H. Griffith. 2005. "Pathological Lying Revisited." *Journal of the American Academy of Psychiatry and the Law Online* 33(3): 342–49.

Doering, Nicola, Katja Richter, Horst-Michael Gross, Christof Schroeter, Steffen Mueller, Michael Volkhardt, Andrea Scheidig, and Klaus Debes. 2016. "Robotic Companions for Older People: A Case Study in the Wild." *Annual Review of Cybertherapy and Telemedicine 2015: Virtual Reality in Healthcare: Medical Simulation and Experiential Interface* 219: 147–52.

Douglas, Thomas. 2008. "Moral Enhancement." *Journal of Applied Philosophy* 25(3): 228–45. https://doi.org/10.1111/j.1468-5930.2008.00412.x

Epictetus. 1983. *The Handbook (The Encheiridion)*. Trans. Nicholas P. White. Indianapolis: Hackett Publishing.

Fischinger, David, Peter Einramhof, Konstantinos Papoutsakis, Walter Wohlkinger, Peter Mayer, Paul Panek, Stefan Hofmann, Tobias Koertner, Astrid Weiss, and Antonis Argyros. 2016. "Hobbit, a Care Robot Supporting Independent Living at Home: First Prototype and Lessons Learned." *Robotics and Autonomous Systems* 75: 60–78. https://doi.org/10.1016/j.robot.2014.09.029

Frank, Lily Eva. 2019. "What Do We Have to Lose? Offloading through Moral Technologies: Moral Struggle and Progress." *Science and Engineering Ethics*, online first. https://doi.org/10.1007/s11948-019-00099-y

Ge, Bi, Hae Won Park, and Ayanna M. Howard. 2016. "Identifying Engagement from Joint Kinematics Data for Robot Therapy Prompt Interventions for Children with Autism Spectrum Disorder." In *International Conference on Social Robotics*, 531–40. Kansas City: Springer. https://doi.org/10.1007/978-3-319-47437-3_52

Germer, Christopher K., Ronald D. Siegel, and Paul R. Fulton. 2016. *Mindfulness and Psychotherapy*. New York: Guilford Publications.

Gibbs, John C. 2010. *Moral Development & Reality: Beyond the Theories of Kohlberg and Hoffman*. Thousand Oaks, CA: Penguin Academics. https://doi.org/10.4135/9781452233604

Gilligan, Carol. 1977. "In a Different Voice: Women›s Conceptions of Self and of Morality." *Harvard Educational Review* 47(4): 481–517. https://doi.org/10.17763/haer.47.4.g6167429416hg5l0

Giubilini, Alberto, and Julian Savulescu. 2018. "The Artificial Moral Advisor. The 'Ideal Observer' Meets Artificial Intelligence." *Philosophy & Technology* 31(2): 169–88. https://doi.org/10.1007/s13347-017-0285-z

Glenn, Andrea L., and Adrian Raine. 2014. "Neurocriminology: Implications for the Punishment, Prediction and Prevention of Criminal Behaviour." *Nature Reviews Neuroscience* 15(1): 54–63. https://doi.org/10.1038/nrn3640

Gordon, John Stewart. 2019. "Building Moral Robots: Ethical Pitfalls and Challenges." *Science and Engineering Ethics*, online first. https://doi.org/10.1007/s11948-019-00084-5

Greene, Joshua D. 2015. "The Rise of Moral Cognition." *Cognition* 135: 39–42. https://doi.org/10.1016/j.cognition.2014.11.018

Gross, Horst-Michael, Steffen Mueller, Christof Schroeter, Michael Volkhardt, Andrea Scheidig, Klaus Debes, Katja Richter, and Nicola Doering. 2015. "Robot Companion for Domestic Health Assistance: Implementation, Test and Case Study under Everyday Conditions in Private Apartments." In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference*, 5992–99. https://doi.org/10.1109/IROS.2015.7354230

Gross, Horst-Michael, Andrea Scheidig, Klaus Debes, Erik Einhorn, Markus Eisenbach, Steffen Mueller, Thomas Schmiedel, Thanh Q. Trinh, Christoph Weinrich, and Tim Wengefeld. 2016. "ROREAS: Robot Coach for Walking and Orientation Training in Clinical Post-Stroke Rehabilitation—Prototype Implementation and Evaluation in Field Trials." *Autonomous Robots* 41(3): 679–98. https://doi.org/10.1007/s10514-016-9552-6

Gupta, Rupa, Daniel Tranel, and Melissa C. Duff. 2012. "Ventromedial Prefrontal Cortex Damage Does Not Impair the Development and Use of Common Ground in Social Interaction: Implications for Cognitive Theory of Mind." *Neuropsychologia* 50(1): 145–52. https://doi.org/10.1016/j.neuropsychologia.2011.11.012

Hamlin, J. Kiley, Karen Wynn, and Paul Bloom. 2010. "Three-Month-Olds Show a Negativity Bias in Their Social Evaluations." *Developmental Science* 13(6): 923–29. https://doi.org/10.1111/j.1467-7687.2010.00951.x

Hirschberg, Julia, and Christopher D. Manning. 2015. "Advances in Natural Language Processing," *Science* 349(6245): 261–66. https://doi.org/10.1126/science.aaa8685

Hoffman, Martin L. 2001. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511805851

Huebner, Bryce. 2015. "Do Emotions Play a Constitutive Role in Moral Cognition?" *Topoi* 34(2): 427–40. https://doi.org/10.1007/s11245-013-9223-6

Hwang, Kwang-Kuo. 2015. "Morality 'East'and 'West': Cultural Concerns." In *International Encyclopedia of the Social & Behavioral Sciences*, ed. James D. Wright, 806–10. Oxford: Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.64005-9

Hyde, Luke W., Daniel S. Shaw, and Ahmad R. Hariri. 2013. "Understanding Youth Antisocial Behavior Using Neuroscience through a Developmental Psychopathology Lens: Review, Integration, and Directions for Research." *Developmental Review* 33(3): 168–223. https://doi.org/10.1016/j.dr.2013.06.001

Irvine, William B. 2008. *A Guide to the Good Life: The Ancient Art of Stoic Joy*. Oxford: Oxford University Press.

Isserow, Jessica. 2015. "Empathy and Morality." *Biology & Philosophy* 30(4): 597–608. https://doi.org/10.1007/s10539-015-9489-8

Jalbrzikowski, Maria, Bart Larsen, Michael N. Hallquist, William Foran, Finnegan Calabro, and Beatriz Luna. 2017. "Development of White Matter Microstructure and Intrinsic Functional Connectivity Between the Amygdala and Ventromedial Prefrontal Cortex: Associations With Anxiety and Depression." *Biological Psychiatry* 82(7): 511–21. https://doi.org/10.1016/j.biopsych.2017.01.008

Jebari, Karim. 2014. "What to Enhance: Behaviour, Emotion Disposition?" *Neuroethics* 7(3): 253–61. https://doi.org/10.1007/s12152-014-9204-5

Johnson, Mark. 2015. *Morality for Humans: Ethical Understanding from the Perspective of Cognitive Science*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226113548.001.0001

Karunarathne, Deneth, Yoichi Morales, Tatsuya Nomura, Takayuki Kanda, and Hiroshi Ishiguro. 2019. "Will Older Adults Accept a Humanoid Robot as a Walking Partner?" *International Journal of Social Robotics* 11(2): 343–58. https://doi.org/10.1007/s12369-018-0503-6

Kazantzis, Nikolaos, Hoang K. Luong, Alexandra S. Usatoff, Tara Impala, Rui Y. Yew, and Stefan G Hofmann. 2018. "The Processes of Cognitive Behavioral Therapy: A Review of Meta-Analyses." *Cognitive Therapy and Research* 42(4): 349–57. https://doi.org/10.1007/s10608-018-9920-y

Kennett, Jeanette, and Cordelia Fine. 2008. "Internalism and the Evidence from Ppsychopaths and 'Acquired Sociopaths.'" *Moral Psychology* 3: 173–90.

Kingma, Elselijn. 2007. "What Is It to Be Healthy?" *Analysis* 67(2): 128–33. https://doi.org/10.1093/analys/67.2.128

Klincewicz, Michał. 2016. "Artificial Intelligence as a Means to Moral Enhancement." *Studies in Logic, Grammar and Rhetoric* 48(1): 171–87. https://doi.org/10.1515/slgr-2016-0061

Klincewicz, Michał, Lily Eva Frank, and Marta Sokólska. 2018. "Drugs and Hugs: Stimulating Moral Dispositions as a Method of Moral Enhancement." *Royal Institute of Philosophy Supplements* 83: 329–50. https://doi.org/10.1017/S1358246118000437

Kluge, Eike-Henner W. 2007. "Resource Allocation in Healthcare: Implications of Models of Medicine as a Profession." *Medscape General Medicine* 9(1): 57.

Kohlberg, Lawrence. 1984. *Essays in Moral Development, vol. 2: The Psychology of Moral Development*. New York: Harper & Row.

Krebs, Denis L., and Kathy Denton. 2005. "Toward a More Pragmatic Approach to Morality: A Critical Evaluation of Kohlberg's Model." *Psychological Review* 112: 629–49. https://doi.org/10.1037/0033-295X.112.3.629

Kumazaki, Hirokazu, Yuichiro Yoshikawa, Yuko Yoshimura, Takashi Ikeda, Chiaki Hasegawa, Daisuke N. Saito, Sara Tomiyama, Jyung-min An, Jiro Shimaya, Hiroshi Ishiguro, Yoshio Matsumoto, Yoshio Minabe, and Mitsuro Kikuchi. 2018. "The Impact of Robotic Intervention on Joint Attention in Children with Autism Spectrum Disorders." *Molecular Autism* 9: 46. https://doi.org/10.1186/s13229-018-0230-8

Lapsley, Daniel, and Gustavo Carlo. 2014. "Moral Development at the Crossroads: New Trends and Possible Futures." *Developmental Psychology* 50(1): 1–7. https://doi.org/10.1037/a0035225

Lapsley, Daniel K., and Darcia Narvaez. 2005. "Moral Psychology at the Crossroads." In *Character Psychology and Character Education*, ed. Daniel K. Lapsley and F. Clark Power, 18–35. Notre Dame, IN: University of Notre Dame Press.

Lara, Francisco, and Jan Deckers. 2019. "Artificial Intelligence as a Socratic Assistant for Moral Enhancement." *Neuroethics*, online first. https://doi.org/10.1007/s12152-019-09401-y

Lehmann, Hagen, Iolanda Iacono, Kerstin Dautenhahn, Patrizia Marti, and Ben Robins. 2014. "Robot Companions for Children with Down Syndrome: A Case Study." *Interaction Studies* 15(1): 99–112. https://doi.org/10.1075/is.15.1.04leh

Leite, Iolanda, Carlos Martinho, and Ana Paiva. 2013. "Social Robots for Long-Term Interaction: A Survey." *International Journal of Social Robotics* 5(2): 291–308. https://doi.org/10.1007/s12369-013-0178-y

Long, Anthony A., and David N. Sedley. 1987. *The Hellenistic Philosophers, vol. 2: Greek and Latin Texts with Notes and Bibliography*. Cambridge: Cambridge University Press.

Magnenat-Thalmann, Nadia, and Zhijun Zhang. 2014. "Assistive Social Robots for People with Special Needs." In *Contemporary Computing and Informatics (IC3I), 2014 International Conference*, 1374–80. The Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/IC3I.2014.7019828

Maibom, Heidi Lene. 2014. *Empathy and Morality*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199969470.001.0001

Motzkin, Julian C., Carissa L. Philippi, Richard C. Wolf, Mustafa K. Baskaya, and Michael Koenigs. 2015. "Ventromedial Prefrontal Cortex Is Critical for the Regulation of Amygdala Activity in Humans." *Biological Psychiatry* 77(3): 276–84. https://doi.org/10.1016/j.biopsych.2014.02.014

Murguia, Edward, and Kim Díaz. 2015. "The Philosophical Foundations of Cognitive Behavioral Therapy: Stoicism, Buddhism, Taoism, and Existentialism." *Journal of Evidence-Based Psychotherapies* 15(1): 37–50.

Murray, Gabriel. 2017. "Stoic Ethics for Artificial Agents." In *Canadian AI 2017: Advances in Artificial Intelligence*, Lecture Notes in Computer Science 10233, ed. Malek Mouhoub and Philippe Langlais, 373–84. Edmonton, AB: Springer. https://doi.org/10.1007/978-3-319-57351-9_42

Narvaez, Darcia, and Daniel Lapsley. 2014. "Becoming a Moral Person: Moral Development and Moral Character Education as a Result of Social Interactions." In *Empirically Informed Ethics: Morality between Facts and Norms*, ed. Markus Christen, Carel van Schaik, Johannes Fischer, Markus Huppenbaur, and Carmen Tanner, 227–38. Cham: Springer. https://doi.org/10.1007/978-3-319-01369-5_13

Padilla-Walker, Laura M. 2014. "Parental Socialization of Prosocial Behavior." In *Prosocial Development: A Multidimensional Approach*, ed. Laura M. Padilla-Walker and Gustavo Carlo, 131–55. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199964772.003.0007

Paulhus, Delroy L., and Kevin M. Williams. 2002. "The Dark Triad of Personality: Narcissism, Machiavellianism, and Psychopathy." *Journal of Research in Personality* 36(6): 556–63. https://doi.org/10.1016/S0092-6566(02)00505-6

Pembroke, Simon G. 1971. "Oikeiosis." In *Problems in Stoicism*, ed. Anthony Arthur Long, 114–49. London: Athlone Press.

Peri, Kathryn, Ngaire Kerse, Elizabeth Broadbent, Chandimal Jayawardena, Tony Kuo, Chandan Datta, Rebecca Stafford, and Bruce MacDonald. 2016. "Lounging with Robots—Social Spaces of Residents in Care: A Comparison Trial." *Australasian Journal on Ageing* 35(1): 1–6. https://doi.org/10.1111/ajag.12201

Persson, Ingmar, and Julian Savulescu. 2012. *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199653645.001.0001

Picard, Rosalind W. 1997. *Affective Computing*. Cambridge, MA: MIT Press. https://doi.org/10.1037/e526112012-054

Pizarro, David A., Brian Detweiler-Bedell, and Paul Bloom. 2006. "The Creativity of Everyday Moral Reasoning." in *Creativity and Reason in Cognitive Development*, ed. James C. Kaufman and John Baer, 81–98. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511606915.006

Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.

Rawls, John. 2009. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Robertson, Donald J. 2016. "The Stoic Influence on Modern Psychotherapy." In *The Routledge Handbook of the Stoic Tradition*, ed. John Sellars, 394–408. London: Routledge.

Robins, Ben, and Kerstin Dautenhahn. 2014. "Tactile Interactions with a Humanoid Robot: Novel Play Scenario Implementations with Children with Autism." *International Journal of Social Robotics* 6(3): 397–415. https://doi.org/10.1007/s12369-014-0228-0

Robinson, Hayley, Bruce MacDonald, and Elizabeth Broadbent. 2014. "The Role of Healthcare Robots for Older People at Home: A Review." *International Journal of Social Robotics* 6(4): 575–91. https://doi.org/10.1007/s12369-014-0242-2

Saghai, Yashar. 2013. "Salvaging the Concept of Nudge." *Journal of Medical Ethics* 39(8): 487–93. https://doi.org/10.1136/medethics-2012-100727

Sandi, Carmen, and József Haller. 2015. "Stress and the Social Brain: Behavioural Effects and Neurobiological Mechanisms." *Nature Reviews Neuroscience* 16(5): 290–304. https://doi.org/10.1038/nrn3918

Savulescu, Julian, and Hannah Maslen. 2015. "Moral Enhancement and Artificial Intelligence: Moral AI?" In *Beyond Artificial Intelligence*, ed. Jan Romportl, Eva Zackova, and Jozef Kelemen, 79–95. Cham: Springer. https://doi.org/10.1007/978-3-319-09668-1_6

Schaefer, G. Owen, and Julian Savulescu. 2019. "Procedural Moral Enhancement." *Neuroethics* 12(1): 73–84. https://doi.org/10.1007/s12152-016-9258-7

Schmidt, Andreas T. 2017. "The Power to Nudge." *American Political Science Review* 11(2): 404–17. https://doi.org/10.1017/S0003055417000028

Schofield, Malcolm. 1995. "Two Stoic Approaches to Justice." In *Justice and Generosity: Studies in Hellenistic Social and Political Philosophy-Proceedings of the Sixth Symposium Hellenisticum*, ed. André Laks and Malcolm Schofield, 191–212. Cambridge: Cambridge University Press.

Schofield, Malcolm. 2003. "Stoic Ethics." In *The Cambridge Companion to the Stoics*, ed. Brad Inwood, 233–56. Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL052177005X.010

Seneca, Lucius Annaeus. 2001. *Epistles: 66–92*. Trans. Richard Mott Gummere. Cambridge, MA: Harvard University Press.

Seneca, Lucius Annaeus. 2004. *On the Shortness of Life, vol. 1*. Trans. Charles Desmond Nuttall Costa. London: Penguin.

Snarey, John R. 1985. "Cross-cultural Universality of Social-Moral Development: A Critical Review of Kohlbergian Research." *Psychological Bulletin* 97(2): 202–32. https://doi.org/10.1037/0033-2909.97.2.202

Snarey, John, and Peter Samuelson. 2008. "Moral Education in the Cognitive Development Tradition: Lawrence Kohlberg's Revolutionary Ideas." *Handbook of Moral and Character Education*, ed. Larry Nucci and Darcia Narvaez, 69–95. New York: Routledge.

Specker, Jona, Farah Focquaert, Kasper Raus, Sigrid Sterckx, and Maartje Schermer. 2014. "The Ethical Desirability of Moral Bioenhancement: A Review of Reasons." *BMC Medical Ethics* 15(1): 67. https://doi.org/10.1186/1472-6939-15-67

Standen, Penny J., David J. Brown, Joseph Hedgecock, Jess Roscoe, Maria Jose Galvez Trigo, and Elmunir Elgajiji. 2016. "Adapting a Humanoid Robot for Use with Children with Profound and Multiple Disabilities." *International Journal of Child Health and Human Development* 9(3): 205–11.

Sterba, James P. 2004. *The Triumph of Practice over Theory in Ethics*. Oxford: Oxford University Press.

Striker, Gisela. 1991. "Following Nature: A Study in Stoic Ethics." *Oxford Studies in Ancient Philosophy* 9: 1–73.

Stueber, Karsten. 2017. "Empathy." In *The Standard Encyclopedia of Philosophy (Spring 2017 Edition)*, ed. Edward N. Zalta. https://plato.stanford.edu/archives/spr2017/entries/empathy/.

Sunstein, Cass R. 2015. "Nudging and Choice Architecture: Ethical Considerations." *Yale Journal on Regulation*. https://ssrn.com/abstract=2551264

Taber-Thomas, Bradley C., Erik W. Asp, Michael Koenigs, Matthew Sutterer, Steven W. Anderson, and Daniel Tranel. 2014. "Arrested Development: Early Prefrontal Lesions Impair the Maturation of Moral Judgement." *Brain* 137(4): 1254–61. https://doi.org/10.1093/brain/awt377

Thaler, Richard, and Cass Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Vaish, Amrisha, Tobias Grossmann, and Amanda Woodward. 2008. "Not All Emotions Are Created Equal: The Negativity Bias in Social-Emotional Development." *Psychological Bulletin* 134(3): 383–403. https://doi.org/10.1037/0033-2909.134.3.383

van Bavel, Jay J., Oriel FeldmanHall, and Peter Mende-Siedlecki. 2015. "The Neuroscience of Moral Cognition: From Dual Processes to Dynamic Systems." *Current Opinion in Psychology* 6: 167–72. https://doi.org/10.1016/j.copsyc.2015.08.009

Walker, Lawrence J. 2004. "Progress and Prospects in the Psychology of Moral Development." *Merrill-Palmer Quarterly* 50(4): 546–57. https://doi.org/10.1353/mpq.2004.0038

Wessells, Michael G. 2006. *Child Soldiers: From Violence to Protection*. Cambridge, MA: Harvard University Press.

Wildschut, Tim, Constantine Sedikides, Jamie Arndt, and Clay Routledge. 2006. "Nostalgia: Content, Triggers, Functions." *Journal of Personality and Social Psychology* 91(5): 975. https://doi.org/10.1037/0022-3514.91.5.975