Why do We Need to Employ Exemplars in Moral Education? Insights from Recent Advances in Research on Artificial Intelligence

Hyemin Han

Educational Psychology Program, University of Alabama

Author Note

Hyemin Han  https://orcid.org/0000-0001-7181-2565

Correspondence concerning this article should be addressed to Hyemin Han, University of Alabama, Box 872031, Tuscaloosa, AL 35487, United States.

Email: hyemin.han@ua.edu

Why do We Need to Employ Exemplars in Moral Education? Insights from Recent Research on Artificial Intelligence

Abstract

In this paper, I examine why moral exemplars are useful and even necessary in moral education despite several critiques from researchers and educators. To support my point, I review recent AI research demonstrating that exemplar-based learning is superior to rule-based learning in model performance in training neural networks, such as large language models. I particularly focus on why education aiming at promoting the development of multifaceted moral functioning can be done effectively by using exemplars, which is similar to exemplar-based learning in AI model training. Furthermore, I discuss the potential limitations and issues related to exemplar-applied moral education with findings from recent studies in AI research raising concerns about model biases and toxic outcomes. I attempt to propose ways to address the concerns regarding employing moral exemplars as well. As remedies, I suggest that autonomy-supporting deliberative and reflective learning processes should be utilized. Furthermore, based on the discussion, I examine how macroscopic socio-cultural aspects influence the effectiveness of exemplar-applied moral education. Suggestions for moral educators and future directions for research in moral education are briefly discussed.

Keywords: Moral education; Moral exemplar; Artificial intelligence; Neural network; Large language models

Introduction

Moral educators have regarded moral exemplars as fundamental educational sources (Kristjánsson, 2006; Sanderse, 2012). Many researchers, including moral

psychologists and philosophers, have studied and supported using exemplars in moral education. First, from the perspective of moral psychologists, psychological evidence supports the practical values of exemplars in moral education. Social psychologists have suggested that presenting one with a moral exemplar, which is seemingly superior, initiates an upward social comparison (Han et al., 2017; Smith, 2000). As a result, the social comparison promotes motivation to do the similar behavior presented by the exemplar to catch up with the perceived gap between the exemplar and themselves (Lockwood et al., 2002; Suls et al., 2002). In addition to social comparison, moral elevation can also explain the mechanism of motivational promotion by moral exemplars (Haidt, 2000). According to Haidt, observing morally superior others may induce a strong positive emotion, i.e., elevation (Haidt, 2000). In previous empirical studies, the perceived elevation after watching moral exemplars promoted motivation for diverse prosocial behavior, such as helping others, donating behavior, etc. (Algoe & Haidt, 2009; Freeman et al., 2009; Schnall et al., 2010; Silvers & Haidt, 2008; Vianello et al., 2010). Finally, the social learning theory proposed by Bandura also explains how exemplars promote moral motivation via vicarious social learning (Bandura et al., 1963). In the moral domain, exemplary conduct can act as an agent to stimulate students' motivation to emulate such conduct to improve their morals and character (Bandura & McDonald, 1963). The moral exemplars work as models for indirect social learning in the moral domain (Bandura et al., 1963; Bandura & McDonald, 1963).

Furthermore, moral philosophers interested in education, particularly those focusing on virtue ethics, have suggested the educational importance of moral exemplars. Their philosophical examinations on the values of exemplars have been widely inspired by

empirical studies focusing on the psychological impacts of diverse exemplars (Athanassoulis, 2022). For instance, Kristjánsson proposed that it is possible to deem emulating[1] role models as an (emotional) virtue, so emulation constitutes the basis of moral education (Kristjánsson, 2006, 2017). Furthermore, based on his point, Henderson further argued that emulation is emotionally, cognitively, and behaviorally virtuous (Henderson, 2024a). She proposed the concept of entangled *phronesis*, practical wisdom. According to this idea, even if students do not possess fully cultivated *phronesis*, which is required for optimally practicing moral virtue, they can habituate *phronesis* by emulating moral role models, such as teachers (Henderson, 2024a, 2023). Hence, emulating moral exemplars is virtuous and a fundamental process towards internalizing and cultivating the virtue of *phronesis* (Athanassoulis & Han, 2023; Han, 2023a).

Several moral philosophical accounts, particularly those related to moral emotions, propose points about moral exemplarity and motivation consistent with the abovementioned psychological research. For instance, some philosophers proposed the potential values of diverse types of exemplars. They suggested that even negative, not positive, emotions originating from seeing superior exemplars can also generate motivation for emulation (Athanassoulis, 2022; Vaccarezza & Niccoli, 2019). Constructive and benign envy perceived after observing one with superior morality and virtue than oneself is supposed to motivate one to emulate such behavior to fill the gap between one and the exemplar (Han et al., 2017; Vaccarezza & Niccoli, 2022). In other instances, when

[1] Emulation within the context of moral education is different from mere imitation (Fridland & Moore, 2015; Kristjánsson, 2006). Mere imitation is about immediately copying a presented behavior, so it does not ensure the implementation of such a behavior across different situations and contexts. On the other hand, emulation is more about learning mechanisms and rules from presented models, develop one's own generalized behavioral strategies, and finally, behave in such a way even in different environments and situations.

one encounters negative exemplars, such as villains, one is likely to feel robust negative emotions so that one will want to avoid the presented morally negative behavior in the future to be not morally worse (Athanassoulis, 2022; Vaccarezza & Niccoli, 2019).

Furthermore, it might also be worth considering the importance of effects from macroscopic socio-cultural factors on moral education, especially exemplar-applied moral education. The Ecological Systems Theory, which has been widely employed in developmental psychology, proposes that researchers should pay attention to not only microscopic factors, such as family and friends, but also macroscopic factors, such as socio-cultural backgrounds and political aspects, while studying developmental trajectories (Bronfenbrenner, 1993). Given this theoretical perspective, I assume that the effectiveness of exemplar-applied moral education, which occurs at the microscopic level, might be significantly influenced by the macroscopic-level factors mentioned above. Murry et al. (2024) also argued that using exemplars in education should be culturally and contextually sensitive to be effective and minimize potential negative outcomes.

Current Paper

The previous philosophical and psychological works have proposed many insights about why and how employing moral exemplars can promote moral motivation, so educators should consider them as fundamental sources for moral education. According to Damon and Colby (2013), moral exemplars are moral paragons who demonstrate moral excellence holistically through all aspects of their lives, including but not limited to cognitive, affective, motivational, and behavioral domains. Likewise, in general, moral exemplars are existential paragons demonstrating virtuous deeds, while role models are more about agents that promote moral emulation among students (Henderson, 2024b).

Despite the values of exemplars, whether the exemplar-applied method might be more effective in moral development than other educational methods still requires further investigation. Several researchers even raise concerns about the validity of the exemplar-applied education in moral education (e.g., Carr, 2023).

Hence, in this paper, I will examine whether it is possible to address the abovementioned question about whether using moral exemplars in moral education is valid and justifiable. For this purpose, I will briefly review recent advances in artificial intelligence in computer science to understand more accurately the mechanism of cognitive and learning processes (Han, 2023b). Of course, some may argue that research on artificial intelligence is not about human cognition per se, so such research cannot significantly inform my investigation of human morality (Frank, 2023; Ke et al., 2024). Thus, I will start by briefly overviewing recent collaborative works in computer science, neuroscience, and cognitive science that showcase why artificial intelligence research can help us better understand human psychology, including the psychology of cognition and learning (Han et al., 2019; Ke et al., 2024; Trott et al., 2023). Based on that, I will review recent studies in artificial intelligence involving exemplar learning and performance improvement to examine why researchers and educators should consider exemplar-applied methods as the core methods in moral education. Finally, I plan to discuss additional educational implications of the exemplary method at diverse levels, from microscopic to macroscopic levels, with the large-scale artificial intelligence research and the model proposed in the Ecological Systems Theory in developmental psychology (Bronfenbrenner, 1993).

AI Research Informs Psychological and Neuroscientific Studies of Human Cognition and Learning

Before starting my discussion on how research on AI can inform our examination of the importance of exemplars in moral education, let me review recent computer science studies in AI along with relevant research in neuroscience and psychology. I plan to conduct this brief review of the technology and its development to discuss whether and how AI research can provide insights into understanding human learning and psychology better. Without such supporting evidence, if there is nothing that AI can inform us about understanding human learning and cognition, I cannot justify the core theme of my paper, learning from AI research to consider the necessity of exemplars in moral education (Frank, 2023; Ke et al., 2024; Trott et al., 2023). If concrete cases demonstrating that AI research can inform human psychology and neuroscience exist, they can support my basic assumption in this paper.

Let me start by briefly overviewing how scientists developed AI at the beginning. When AI research became a hot topic in computer science, computer scientists attempted to reverse-engineer human psychology and neuroscience to create artificial neural networks that can learn from environmental factors and produce outcomes like humans (Redford et al., 1995; Saxe, 2015). For instance, the deep learning technology currently widely utilized for prediction, classification, and other computational tasks employs multiple layers of neural networks consisting of artificial neurons (LeCun et al., 2015). Computer scientists designed artificial neurons to generate outputs based on input values similar to human neurons. A multi-layer network of artificial neurons makes predictions based on external input like human brains do (Macpherson et al., 2021). Then, feedback for

reinforcement originating from the difference between the prediction and reality, adjusts weights of the artificial neurons to improve prediction performance continuously (Saxe, 2015).

Likewise, large language models (LLMs) that process natural language inputs and generate predictive outcomes based on large-scale statistical modeling also simulate human learning and predictive processes (Ke et al., 2024; Zhao et al., 2023). LLMs learn patterns from large-scale human language datasets with provided feedback and labeling information (Chang et al., 2023). Once the model training procedures are complete, they can generate outputs with the highest predictive likelihoods based on input queries in natural language (Han, 2023b; Zhao et al., 2023). As a result, they successfully respond to inquiries in diverse domains, including but not limited to reasoning, social cognition, and clinical sciences, like humans or even domain experts (Kasneci et al., 2023; Zhao et al., 2023). For example, several recent studies employing LLMs have demonstrated they can address sophisticated matters relevant to morality, such as generating philosophical accounts and simulating the theory of mind and perspective-taking (Chalmers, 2023; Kosinski, 2023; Schwitzgebel et al., 2023; Williams et al., 2022). Thus, researchers in psychological science are utilizing LLMs in their research projects frequently for various reasons, such as modeling human cognition and behavior and simulating human subjects (Demszky et al., 2023; Han, 2023b; Ke et al., 2024).

Now, scientists try to employ AI systems developed originally by simulating human psychological and neural processes to understand human psychology and neuroscience more accurately (Koppe et al., 2021; Macpherson et al., 2021). Researchers are utilizing predictive technologies based on artificial intelligence, such as deep learning technology, to

examine factors predicting psychological and behavioral outcomes (e.g., Han et al., 2020). They also investigate the psychological mechanisms in the processes by looking into the trained artificial neural networks (Demszky et al., 2023; Macpherson et al., 2021). For instance, in moral psychology, one preliminary study employing LLMs suggested that LLMs can address ethical dilemmas and learn from the stories of exemplars like human participants (Han, 2023b). The author also suggested that LLMs will help moral psychologists by enabling them to simulate psychological processes in moral domains without any risk involving human participants. At the neural level, neuroscientists are now applying artificial neural networks to model and simulate human brain activities (Macpherson et al., 2021; Rashid et al., 2020). For example, we may consider recent neuroscientific studies that examined the neural correlates of visual processing and abnormal neural networks among patients with schizophrenia (e.g., Alves et al., 2023). In these studies, researchers employed artificial neural networks to simulate and model accurately the neural correlates of interest.

Given these examples at the behavioral, psychological, and neural levels, it is the case that research on AI can significantly inform our examinations of human psychology and neuroscience (Chang et al., 2023; Han, 2023b; Macpherson et al., 2021). As mentioned above, investigating artificial neural networks and prediction models simulating human cognition and learning may provide insights into how to understand them with methodological efficiency. Consequently, my basic assumption in the present paper, i.e., we will be able to learn about the psychological processes involving moral learning from AI research, is assumably supported by the overviewed research trends.

Despite the similarities in cognitive processes between human and AI that have been discussed, some may still argue that AI systems are fundamentally different from human brains in terms of how they are implemented and operated (e.g., Chinese Room argument) (see Cole, 2020). Based on that, they may also criticize the validity of the analogy that I introduced (e.g., Giannakidou & Mari, 2024). I understand that it might be too early to conclude that AIs are completely identical to human brains in terms of their functioning and implementation. However, I assume that the analogy still has several practical benefits within the context of the current study (see Lengbeyer [2022] and van Dijk et al. [2023] for further discussions regarding supporting the pragmatic values of considering AI research in research on human cognition).

One major concern is that human cognition and psychological processes involving moral functioning are complicated than what have been implemented by AI or LLMs (see Aharoni et at., 2024 for overview). Although I agree this point, I shall argue that such a difference between AI and human, particularly, the difference in functional complexity, cannot significantly weaken the basis of my point about the utility of exemplar-applied education. In the following section, I will overview the previous AI studies demonstrating that simply teaching rules and focusing on specific capacities are not effective in training complicated cognitive processes (Hadi et al., 2023; Yamazaki et al., 2023). Instead, AI research has shown that using concrete exemplars is the best way to develop capacities for cognitive process and complicated problem-solving. If this is the case among AIs, then human moral functioning, which requires more capacities and complicated cognitive processes, could also not be developed via simple teaching methods without employing real exemplars. Perhaps, moral education for human beings even more strongly requires

exemplars than moral learning for AIs due to the greater complexity within their cognitive and moral functioning. Hence, although the concern about the functional and structural differences between AIs and human brains per se might be valid from the conceptual perspective, such a concern would not significantly threat the plausibility of my proposal based on the abovementioned analogy from the practical and pragmatic perspective.

In this section, I reviewed previous research on AI to propose: first, computer scientists developed and validated artificial neural networks to implement AI, such as deep learning and LLMs, by analyzing and simulating human brains and psychological processes (LeCun et al., 2015; Macpherson et al., 2021; Saxe, 2015); and second, now neuroscientists and psychologists are examining and employing AI technologies to understand the mechanisms of human cognition and learning at the behavioral, psychological, and neural levels (Alves et al., 2023; Han, 2023b; Ke et al., 2024; Koppe et al., 2021; Macpherson et al., 2021). In conclusion, these might support my point that examining AI will provide insights into whether and how studying AI will inform our consideration of using exemplars and role models in moral education. Although there might be some concerns due to the functional and structural differences between cognitive processes implemented in AIs and humans as I overviewed, my point about the necessity of using exemplars in moral education can still be supported from the practical perspective. Further details about this point will be discussed in the following section.

Exemplar-based Learning for Developing Multifaceted Moral Functioning

In this section, I will discuss why employing exemplars can be a feasible way to promote moral development in moral education. While addressing this topic, I will draw upon theoretical and empirical developments in AI research, particularly recent works

using example-based AI models, to support my point. Finally, I will discuss why my point can still be supported despite differences in the learning mechanisms between AIs and humans.

Previously, computer scientists developed AI models based on rules, e.g., combinations of if-then clauses (Campolo & Schwerzmann, 2023). For instance, in an illustrative case of a clinical rule-based AI model, we may imagine a model generating predictive diagnoses based on input data (Van Der Waa et al., 2021). Engineers may create the rule-based AI model following how human clinicians and practitioners render prescriptions with patient stories, test results, etc (Yamazaki et al., 2023). The AI model makes a diagnosis decision following specific rules and criteria entered and programmed previously (e.g., diabetes: if the blood sugar level is higher than…, if the patient lost weight, etc.) (Campolo & Schwerzmann, 2023; Van Der Waa et al., 2021).

In many cases, when the presented problems are well classified and defined, rule-based AI models demonstrate acceptable prediction power and reliability (Wang et al., 2023). However, the rule-based AI models perform worse when the problems they encounter do not fall into the realm of what the pre-learned rules and criteria can address directly (Yamazaki et al., 2023). In other words, rule-based AI lacks generalizability to deal with various real problems, which might be irregular (Hadi et al., 2023; Wang et al., 2023). In such cases, AI cannot appropriately evaluate the problems merely with determined rules and criteria.

Researchers have developed example-based AI, which constitutes the basis for the state-of-art AI models, as an alternative to address the limitations of rule-based AI (Hadi et al., 2023; Yamazaki et al., 2023). As I briefly described in the previous section introducing

neural networks and LLMs, example-based AI models learn patterns from examples for prediction. Instead of explicitly setting rules and criteria for decision-making like rule-based AI, exemplar-based AI models adjust and refine their prediction models, such as those constituted by artificial neural networks, based on input and feedback (Hadi et al., 2023).

Because exemplar-based AI spontaneously learns patterns for prediction, it has enhanced flexibility to address various real problems (Hadi et al., 2023). Even when a rendered prediction is sub-optimal, exemplar-based AI can adjust its prediction models via reinforcement and feedback to improve prediction power (Ganguli et al., 2023; Schramowski et al., 2022). In reality, studies demonstrate that exemplar-based AI significantly outperforms rule-based AI in diverse domains, including but not limited to clinical diagnosis, natural language processing, conversational interaction, hospitality dialog system, etc., in terms of prediction accuracy, ability to solve complex problems, and flexibility (Chen et al., 2022; Hadi et al., 2023; Wang et al., 2023; Yamazaki et al., 2023). When researchers provided AI models with training data consisting of diverse contextual information, the models demonstrated superior predictive performance compared to when they received rather abstract training data without contextual information (Liévin et al., 2023; Singhal et al., 2022). Furthermore, Nolfi proposed that AI models trained by examples, such as LLMs based on large-scale language corpora, can acquire unexpected cognitive abilities, which are out of the scope of the corpora per se (Nolfi, 2023). They have flexibility in learning to develop generalizable cognitive skills and capabilities indirectly from presented examples.

Moral educators may consider a similar point to improve methods for moral education. As I discussed in the previous section, given research on AI models can provide us with insights into understanding human cognition and learning more accurately (Macpherson et al., 2021), examining rule-based versus example-based AI will also help us address our concern about moral education. Educational methods aiming at improving specific sets of moral functioning, such as moral judgment and reasoning, might be less suitable than exemplar-applied methods for potential generalizability (Carr, 2023; Penn, 1990). For instance, let us imagine that one moral educator is about to teach skills for ethical decision-making through non-exemplar-applied ways. If the educator intends to rely strictly on rule-based methods, the decision-making skills taught in the class could not be well generalizable and applicable in situations other than what students explicitly address in the classroom (e.g., a set of if-then clauses, such as "if innocent victims are about to be harmed, then it shall not be ethically acceptable," etc.) (see Han, 2015). One may argue that conventional dilemma-involved methods, such as dilemma discussions, might be counterexamples demonstrating that non-exemplar-applied methods can be effective (Carr, 2023). However, as proposed by classical Kohlbergians, even dilemma debates in classrooms become ways to present exemplary moral reasoning, such as "plus one," to promote students' development (Bandura & McDonald, 1963; Blatt & Kohlberg, 1975). Hence, it might be possible to conclude that strict rule-based educational activities can hardly develop students' moral functioning in diverse domains.

Instead, exemplars and role models in the moral domain can be more effective sources for moral education and development, as shown by AI research. Concrete exemplars will provide students with context-rich inputs to facilitate moral development

across diverse functional domains (Athanassoulis & Han, 2023; Han, 2023a, 2024a). As Damon and Colby (2013) proposed, moral exemplars are moral paragons who demonstrate moral excellence holistically through all aspects of their lives across all individual capacities for moral functioning. Given it is impossible to explain the mechanism of optimal moral functioning only with a limited number of individual constructs (Darnell et al., 2019; Kristjánsson & Fowers, 2022), for effective moral education and development, presenting exemplars and role models who show holistically superior performance with concrete contextual information might be required (Athanassoulis & Han, 2023; Han, 2023a, 2024a). Hence, via indirect and unexpected learning and the generalization of learned capacities (Dhar, 2023; Singhal et al., 2022; Trott et al., 2023; Van Der Waa et al., 2021; Yamazaki et al., 2023), exemplar-based approaches will offer unique benefits in moral education, similar to the case of example-based learning in AI.

One caveat is that, as I briefly discussed at the end of the section examining the similarities between AIs and humans, AI agents learn patterns via simple statistical learning while human students learn from exemplars via not only statistical adjustment but also emotional processes (Hadi et al., 2023), such as admiration (Zagzebski, 2013). Although I admit that there might be fundamental differences between them, I shall suggest that such a point even further supports the necessity of exemplar-applied learning in moral education from the practical perspective (see for Lengbeyer [2022] and van Dijk et al. [2023] further discussions). Again, it might be the case that human moral development is a significantly more complicated process than statistical adjustment done by AI agents (Aharoni et al., 2024). If so, such complexity may further strengthen the point that moral development could not be done by rule-based learning or educational activities focusing on

one specific functionality. Moral exemplars might be the only effective educational materials that can present the complicated aspects of moral functioning that engage both reasoning and emotion (Damon & Colby, 2013).

Some may also argue that employing AI as an exemplar is unnecessary to support the importance of exemplars in moral education. They may consider that the complexity of the factors influencing developmental trajectories in real world per se can be sufficient evidence substantiating the importance. Such a point seems to be like the critiques to educational neuroscience, which underscored the employment of neuroscience in improving education, due to redundancy (e.g., Dougherty & Robey, 2018). However, I assume that the AI case, which demonstrates the superiority of the exemplar-based learning to the rule-based learning can make a unique contribution to the current work. Previous AI research compared those two learning methods can provide concrete empirical evidence at the infrastructure level (Han et al., 2019) presenting why exemplars should be utilized instead of mere focusing on rules or individual capacities to address complicated developmental processes effectively. By examining the performance of those methods in various problem domains in reality (Hadi et al., 2023; Yamazaki et al., 2023), AI research can contribute to the generalized assumption that concrete exemplars should be used for effective learning.

In the next section, we will refer to research on *phronesis*, practical wisdom, in moral philosophy and education to examine further details about why exemplar-based approaches are necessary for holistic and integrative moral development for optimal moral functioning (Darnell et al., 2019, 2022; Kristjánsson et al., 2021). Given moral philosophers and psychologists argue that *phronesis* is a complex, multifaceted entity and process

involving different psychological functionalities (Darnell et al., 2022; De Caro et al., 2018; Han, 2024b), it would be a great example suggesting why example-based learning is superior to rule-based learning in moral education.

Phronesis Cultivation as a Concrete Example Suggesting Why Moral Education Requires Example-based Learning

In recent research on moral functioning in moral philosophy and psychology, many researchers propose that *phronesis* is fundamental in enabling one to render optimal ethical decisions across different situations (Darnell et al., 2019; De Caro et al., 2018; Kristjánsson & Fowers, 2022). Several studies have demonstrated that this construct successfully predicts moral motivation and behavior, so it is worth researchers' and educators' attention (Darnell et al., 2022; Han, 2024b). Related to the conceptual complexity, recent conceptual examinations of *phronesis* propose several models to explain its organization and mechanism (Han, 2024a; Vaccarezza et al., 2023). Currently, two most representative *phronesis* models exist: the Jubilee Centre Model (Kristjánsson & Fowers, 2022) and the Aretai Centre Model (Vaccarezza et al., 2023).

First, according to the Jubilee Centre Model, *phronesis* is a multifaceted entity consisting of four psychological capacities, i.e., the blueprint of flourishing, moral sensitivity, reason-infused emotion, and moral adjudication[2] (Darnell et al., 2019; Kristjánsson & Fowers, 2022). Optimal moral functioning occurs when these four successfully coordinate and cooperate to achieve the ultimate goal, i.e., flourishing (Hacker-

[2] The blueprint of flourishing is about the extent to which one understands the importance of flourishing and how such understanding motivates and directs their action. Moral sensitivity is related to one's ability to detect morally salient aspects of a situation and figure out the best solution. Reason-infused emotion deals with regulating one's emotions for optimal emotional experiences with guidance of reasoning., Moral adjudication is about balancing different virtues and strengths to be able to render a best decision in a situation while addressing conflicts.

Wright, 2023; Kristjánsson et al., 2021). Cultivation of *phronesis* consequently requires the development of these capacities (Kristjánsson & Fowers, 2022). Moreover, the Aretai Centre Model proposes that *phronesis* implies possessing ethical expertise to exercise individual virtues and strengths in an appropriate manner to address situational concerns (De Caro et al., 2021; Vaccarezza et al., 2023). From their perspective, *phronesis* is an integrative and holistic expertise to produce optimal motivational and behavioral outcomes across different situations. Accordingly, cultivating *phronesis* requires mastering the expertise, probably as practical unity (Hacker-Wright, 2023). As an integrative view, Han (Han, 2024a) proposes that the two models explain two aspects of *phronesis*, i.e., the multifacetedness of its organization (corresponding to the Jubilee Centre Model) and the network-nature in its functioning (corresponding to the Aretai Centre Model). He argues that evidence from psychology and neuroscience supports such an integrative explanation of the nature of *phronesis*.

We also need to consider that the adaptive adjustment of learning processes occurs during *phronesis* development. *Phronesis* cultivation also requires adaptive adjustment of the learning process in moral learning, which is similar to the adjustment of learning parameters in AI training (FeldmanHall & Lamba, 2023; Han, 2023a). Moral philosophers proposed that *phronesis* cultivation requires reflections and deliberations upon experiences (Kristjánsson & Fowers, 2022). Depending on different situations, one may need to adjust existing beliefs and values significantly or maintain them for better decision-making and moral behavior in the future (FeldmanHall & Lamba, 2023; Han, 2023a). For instance, if one encounters dilemmas that challenge current conventions and norms, such as proposals for human rights, one will need to change their views significantly. When one

sees extremist propaganda, conversely, which will severely threaten people's welfare, then such input should not move their beliefs. Only careful post-experiential reflection and deliberation can enable the optimal adjustment of the learning process. The process is very complicated, so approaches merely targeting specific cognitive capacities cannot effectively develop students' abilities to adjust their learning process optimally for *phronesis* cultivation (Athanassoulis & Han, 2023; Han, 2023a, 2024a).

As shown in the standard models and discussed by researchers, it is clear that *phronesis* as an agent rendering optimal moral decisions is sophisticated and complicated in terms of its structure and mechanism (Darnell et al., 2019; Han, 2024a; Kristjánsson & Fowers, 2022). It consists of multiple functional components (as proposed in the Jubilee Centre model) (Darnell et al., 2022) while playing its roles integratively (as proposed in the Aretai Centre model) via networking (De Caro et al., 2021; Hacker-Wright, 2023; Han, 2024a; Vaccarezza et al., 2023). If that is the case, merely teaching ethical rules and individual skill components constituting *phronesis* cannot effectively cultivate *phronesis*. Instead, presenting *phronesis* exemplars possessing developed individual functional components and an ability to coordinate them appropriately across different situations might be the best sources promoting students' *phronesis* development (Han, 2023a, 2024a; Kristjánsson & Fowers, 2022).

Let us briefly review the decathlon analogy employed by Kristjánsson and Fowers (2022) to support the multifaceted and integrated nature of *phronesis* and a *phronesis* expert. They used an example of a professional decathlon athlete to describe the aspects of a *phronesis* expert. The decathlon professional should know how to distribute a limited amount of energy to different components and develop one's physical abilities and skills

integratively to perform well overall. Merely developing skills for individual sports is not sufficient to succeed. It is similar to the case of *phronesis*, which requires the integration of different skill sets and forms an optimal functional network among them. Given these, to train decathlon athletes effectively, merely teaching skills for individual sports (similar to the case of rule-based learning) cannot be appropriate. Instead, only successful decathlon professionals can give them advice on how to coordinate and adjust different components to produce the optimal performance. They can also be great mentors to share their experiences with fellow athletes and discuss their concrete concerns and questions as exemplars.

We may also apply the same point to the *phronesis* exemplars to support exemplar-based, not rule-based, education for *phronesis* cultivation. Given *phronesis* is not merely about whether one can exercise individual virtues or strengths (Hacker-Wright, 2023; Han, 2015; Kristjánsson et al., 2021), simply training such individual capacities does not necessarily and sufficiently cultivate *phronesis*. Like the case of the decathlon exemplar, a *phronesis* exemplar can demonstrate well-balanced, optimal moral functioning to students as a concrete example (Kristjánsson & Fowers, 2022). They can show how to coordinate individual functional components in the network for moral functioning appropriately across different situations. Also, they can provide information about adjusting the learning process when students encounter different external situations challenging their existing views and beliefs (Han, 2023a). *Phronesis* exemplars can facilitate optimal moral learning as living mentors for self-cultivation and autonomous moral growth (Athanassoulis & Han, 2023; Henderson, 2023). Finally, because *phronesis* requires one's capacity to deal with diverse situations in reality (Hacker-Wright, 2023; Kristjánsson, 2014), which should be

well generalizable, an exemplar-based approach might be a viable educational approach, as shown by AI studies (Campolo & Schwerzmann, 2023; Van Der Waa et al., 2021). Rule-based learning focusing on specific skill sets would not suffice such requirements for *phronesis* cultivation (Singhal et al., 2022; Wang et al., 2023; Yamazaki et al., 2023).

Like the case of general exemplar-based moral education that I discussed in the previous section, one caveat is that human virtue acquisition and *phronesis* cultivation might be significantly more complicated than the statistical adjustments done by AI models (Aharoni, 2024; Kristjánsson & Fowers, 2022). For instance, virtue and phronesis development are likely to require cultivation in additional aspects, such as emotion, which has not been fully implemented in AI (Hadi et al., 2023). Hence, some may argue that AI's capacities to adjust its model via example-based learning are significantly different from human's capacities to acquire virtue and cultivate *phronesis*, so my attempt to connect those two within the context of moral education might not be plausible. Even if that is the case, such a fact may support the necessity of using exemplars in virtue education. If virtue and *phronesis* cultivation are complicated processes that could not be fully achieved via simple statistical adjustment, rule-based learning and activities focusing on individual capacities could not be effective educational solutions. Instead, virtuous exemplars, who demonstrate how to perform optimally while coordinating various capacities within a complicated functional network, shall be effective sources for moral education (see Han, 2023a).

In conclusion, exemplars and role models are essential in moral education, given what AI research has demonstrated. Rule-based learning directly and explicitly focusing on specific skill sets cannot produce generalizable outcomes out of the boundary of training.

On the other hand, example-based learning, which constitutes the basis for modern AI models, can train one to use the learned capacities across diverse domains and situations. Such an approach also effectively promotes the development of various abilities and skills through indirect and unexpected learning. Because moral functioning requires the coordination and adjustment of multiple functional components, as shown by *phronesis*, moral educators need to employ exemplar-based approaches as foundational educational methods. That might be the way to address the complexity of moral functioning and generalizability issues involved in moral education.

Some Concerns Regarding the Use of Example-based Approached in Moral Education: Overview of Potential Issues and Future Directions

I will briefly discuss potential challenges regarding using exemplars in moral education in light of recent critiques of AI technologies, particularly LLMs. As shown by AI research that underscores the importance of examples for effective learning (Campolo & Schwerzmann, 2023; Hadi et al., 2023; Van Der Waa et al., 2021; Yamazaki et al., 2023), using moral exemplars and role models is essential in moral education. That is particularly the case if moral educators intend to promote the development of moral functioning across diverse situations in reality, such as *phronesis* (Athanassoulis & Han, 2023; Han, 2024a). However, recent critiques of AI models, particularly LLMs, in socio-cultural contexts raise concerns about the reliability and validity of example-based learning (Navigli et al., 2023; Schramowski et al., 2022; Shaikh et al., 2023). I will first discuss such concerns based on Bronfenbrenner's Ecological Systems Theory (Bronfenbrenner, 1993). Then, I will consider future directions to address the potential issues in educational practice.

Many people, including non-experts in computer science, have found that LLMs, one of the most widely used AI models in recent days, are very useful and versatile in addressing their questions in various domains (Zhao et al., 2023). However, simultaneously, many users and researchers raise concerns about potential biases embedded in predictions and responses made by LLMs (see a consensus paper, Srivastava et al., 2023). Given developers train and tune LLMs with large-scale language datasets collected around the world, LLMs may likely generate biased responses based on diverse forms of biases existing in reality (Navigli et al., 2023; Srivastava et al., 2023). For example, recent studies have reported that LLMs tend to make predictions that are biased in terms of race, ethnicity, religious beliefs, and other forms of sociocultural backgrounds (Abid et al., 2021; Naous et al., 2023; Navigli et al., 2023). One concrete example shows that LLMs are likely to imitate and reproduce hate and discriminatory speeches that are already prevalent within a society (Urman & Makhortykh, 2023). That said, example inputs from macroscopic-level environments around the world surrounding a system can significantly influence the reliability and credibility of outputs generated by the system, possibly negatively (Srivastava et al., 2023).

We may also examine such a concern in the case of exemplar-applied education among human students. If we only consider a single exemplar-student relationship, the abovementioned bias could not be a significant issue. However, like LLMs, numerous entities existing in the world other than the exemplar influence the formation of students' moral beliefs and values (Engelen et al., 2018). I will refer to the Ecological Systems Theory briefly to examine why that is the case and can be a potential problem. According to the Ecological Systems Theory, different systems at different levels, including the microsystem,

mesosystem, and macrosystem, influence one's developmental trajectories (Bronfenbrenner, 1993). For instance, friends, family members, and teachers, who are directly interacting with an individual and most likely to be influential relatable exemplars, are within the realm of the microsystem (Han & Dawson, 2023; Han & Graham, 2023; Lockwood & Kunda, 1997; Piel et al., 2017). The macrosystem is a system surrounding the systems at the lower levels, such as a culture, country, or society (Forbes et al., 2022). Cultural norms, conventions, and social and political policies are the agents that exert influence on an individual from the macrosystem (Nartey et al., 2023). Finally, the mesosystem deals with interconnections and interactions between the major systems mentioned above in individuals' lives (Cowan & Swearer, 2004).

According to the Ecological Systems Theory, individual exemplars in the microsystem are not the only agents influencing one's motivation, although they are most direct, relatable, and influential. That suggests we should also consider how large-scale examples in the moral domain (macro system) influence moral development and how individual exemplars interact with different systems (mesosystem) to understand the example-based learning process (Bronfenbrenner, 1993; Nartey et al., 2023; Piel et al., 2017). I will discuss these two points with some illustrative examples.

First, let us briefly consider the case of large-scale influences of the macrosystem. For example, we may imagine a situation where a student lives with virtuous close exemplars, such as friends, family members, teachers, and community members who implement virtues and moral values (Henderson, 2024a, 2023; Kristjánsson, 2006; Sanderse, 2012). In such a case, it is predictable that the influential relatable exemplars positively influence the student's moral motivation and development, as suggested by

previous moral exemplar studies (e.g., Čehajić-Clancy & Bilewicz, 2021; Han et al., 2017, 2022; Han & Dawson, 2023; Han & Graham, 2023). However, we can consider a society where large-scale socio-cultural norms do not value virtuous actions (e.g., in the case of a totalistic society that promotes unethical extremism) as a counterexample (Niebuhr, 2013). Due to the negative influences from the macrosystem, the individual exemplars at the microscopic level might only be able to produce limited positive influences.

Such a situation is conceptually equivalent to when LLMs learn biased examples from large-scale datasets from society and generate ethically problematic predictions, such as the generation of hate and discriminatory speeches based on the probabilistic prediction of the prevalence of such expressions in the real world (Abid et al., 2021; Naous et al., 2023; Navigli et al., 2023; Srivastava et al., 2023). When the existing reality at the large scale is of suboptimal quality, even if we attempt to train an exemplar-based AI model with appropriate examples at the local level (Navigli et al., 2023, 2023; Srivastava et al., 2023), the model might not be able to demonstrate optimal performance like the case of the student surrounded by individual moral exemplars living in an anti-moral society (Niebuhr, 2013). For instance, as shown in the concrete example introduced previously (e.g., Urman & Makhortykh, 2023), if discrimination against specific minority groups of people is prevalent in one society, then LLMs trained in such a context are likely to reproduce the prevalent hate and discriminatory speeches. In such a case, one's effort to correct such implicit hate and discrimination existing in the LLMs at the local level is less likely to be successful.

Second, we should also carefully consider the interactive influences of the higher-level systems, including the mesosystem. As mentioned above, continuous interactions

between different systems at different layers also affect one's development. In the case of exemplar-based learning, such a point in the higher-level systems may imply that how society, culture, and other agents treat individual exemplars and how individual systems evaluate each other's exemplarity significantly determine the motivational impacts of the exemplars (Bandura & McDonald, 1963; Engelen et al., 2018). For instance, we can imagine a society that highly values morality and honors individual exemplars who implement such values and virtues. In that society, the socio-cultural atmosphere at the macroscopic level and individual systems honor such significantly boost the positive motivational impacts of moral exemplars (Čehajić-Clancy & Bilewicz, 2020; Niebuhr, 2013). Contrarily, if people in a society do not highly regard ethical values and they pursue non-moral values while compromising moral values, such a socio-cultural atmosphere may diminish the voices and messages of moral exemplars (Čehajić-Clancy & Bilewicz, 2020; Hamilton, 2019).

Likewise, in the cases of exemplar-based AI models, recent advances in research on meta-learning in AI[3] may suggest why that can be problematic, as shown by the illustrative example above (Binz et al., 2023). Meta-leaning, which means learning about learning, explains how one AI model can train patterns and perform predictions across different task domains (Binz et al., 2023; Langdon et al., 2022). During the meta-learning process, feedback from outside of one specific domain at the global level can reinforce how an AI model in that domain performs predictions. For instance, in the case of the Theory of Mind (ToM) research, an AI model implementing meta-learning was able to transfer information from learning about the mental state of a specific agent (A) to the case of another agent (B).

[3] I plan to discuss further details about the implications of meta-learning in AI research in moral psychology and moral education. However, that is out of the scope of the current paper, so that will be addressed in another paper.

The trained ToM to infer A's mental status was utilized as prior information to initiate learning about B's mental status effectively thanks to the presence of the global-level meta-learning mechanism (Rabinowitz et al., 2018). Hence, biased example input from the external domains can eventually bias individual models despite appropriate example-applied learning at the local level (Binz et al., 2023; Navigli et al., 2023, 2023). Such a mechanism of meta-learning in AI may support the point that influences from the higher-level systems also significantly affect or even disrupt learning from individual exemplars (Bandura & McDonald, 1963; Čehajić-Clancy & Bilewicz, 2020; Engelen et al., 2018; Niebuhr, 2013).

Since potential negative influences from the macro and mesosystems, as shown with the recent AI research, can be significant concerns, we may need to consider how to address such concerns. Furthermore, such issues can even become opportunities to promote moral development further via exemplar-applied education. First, we can consider the value of supervision during the learning process (Ganguli et al., 2023; Schramowski et al., 2022). Although researchers are concerned about the bias in the trained LLMs due to suboptimal large-scale training datasets (Abid et al., 2021; Navigli et al., 2023; Srivastava et al., 2023), LLMs demonstrate that they can correct their models by receiving appropriate feedback and supervision (Ganguli et al., 2023; Navigli et al., 2023; Schramowski et al., 2022). Even in the moral domain, LLMs could address their antimoral toxic responses, such as hate and discriminatory speeches in the real world, to inquiries emerging from human-like biases with guidance provided as feedback (Ganguli et al., 2023).

Such findings on the necessity of appropriate guidance and reinforcement to address anti-moral biases in LLMs in AI research provide insights into utilizing appropriate

educational approaches, such as self-cultivation and moral mentorship, while applying moral exemplars (Athanassoulis & Han, 2023; Han & Graham, 2023; Sanderse, 2023). As suggested in previous exemplar studies, merely presenting moral exemplars to students could not be an ideal educational approach (Athanassoulis & Han, 2023; Carr, 2023). Without additional guidance, students might be exposed to biased examples during the learning process, such as widely available discrimination against minority groups, like the case of the abovementioned LLMs. Recently, Sanderse (2023) argued that moral educators should employ more self-oriented, self-cultivating educational approaches while using ethical role models. Athanassoulis and Han (2023) also suggested moral exemplars should be mentors providing students with concrete advice on moral matters. Thus, as proposed by these researchers, moral educators should carefully consider employing diverse educational methods for role modeling, particularly those involving interactive activities for self-cultivation, instead of unidirectional story presentation (Athanassoulis & Han, 2023; Han & Graham, 2023; Henderson, 2023; Sanderse, 2023).

Second, AI research on the chain of thought (CoT) (Wei et al., 2023), which addresses the reasoning-based approach to improve AI's performance (Li et al., 2023), suggests employing reflective, deliberative, and autonomous activities during exemplar-applied education. Recent studies examining the performance of diverse AI models, particularly LLMs, demonstrated that when one presents an AI model with the CoT consisting of examples containing the reasoning process involving problem-solving[4] (Ho et al., 2023; Prystawski et al., 2023), the model reported significantly better performance than

[4]E.g., "Corgi has six squeeky puppets. She buys three more boxes of puppets. Each box has two puppets. How many squeeky puppets does she have now?" "**Corgi started with six squeeky puppets. Three boxes of two puppets each is six puppets. 6 + 6 = 12.** The answer is 12."

when one enters explicit examples[5] (Mu et al., 2023; Wei et al., 2023). That said, training AI models is more effective when the models were presented with the actual reasoning processes rather than the mere example outcomes per se in terms of prediction accuracy and quality. The positive effect of the CoT was further improved when one provided additional contextual information along with reasoning processes (Liévin et al., 2023). One point related to moral issues is that depending on guidance and directions provided by developers, the CoT can exacerbate anti-moral biases in trained models (Shaikh et al., 2023), or applying the CoT can facilitate the detection and mitigation of such biases (e.g., why and how the current case can be morally problematic or justifiable) (Ganguli et al., 2023; Tian et al., 2023).

The improved performance of AI with the CoT and the necessity to address anti-morality and biases in trained models while employing the CoT suggests that moral educators should also consider applying a similar approach while using moral exemplars and role models in moral education. Consistent with what we discussed previously, we may apply the benefits of role modeling with self-cultivation to this point (Athanassoulis & Han, 2023; Han, 2023a; Han & Graham, 2023; Sanderse, 2023). By actively and autonomously evaluating, deliberating, and discussing the values and implications of presented moral exemplars, it is possible to promote students' moral motivation more effectively. During the course, students can also critically reflect upon the potential negative influences from the meso- and macrosystem, along with those of individual exemplars. Athanassoulis

[5] E.g., "Corgi has six squeeky puppets. She buys three more boxes of puppets. Each box has two puppets. How many squeeky puppets does she have now?" " **The answer is 12.**"

(Athanassoulis, 2022) suggested even negative figures can morally inspire students via appropriate educational approaches.

Also, biases at the sociocultural level, which may act as negative examples, will become sources for further contextually sensitive moral growth through activities involving deliberative and reasoning processes (Murry et al., 2024). In terms of methodological autonomy, a recent data synthesis revealed that the autonomous aspect of such educational approaches during role modeling is fundamental in promoting moral motivation (Han & Graham, 2023). Likewise, autonomously exercising reasoning skills during exemplar-applied education can contribute to *phronesis* cultivation (Athanassoulis & Han, 2023; Henderson, 2024a, 2023). For example, moral educators and mentors will be able to utilize negative exemplars existing in society who involve culturally and contextually insensitive behavior (e.g., hate speech and discrimination) as sources for autonomous and spontaneous moral deliberation and reflection to promote moral wisdom among students. Given these, autonomous educational activities stimulating students' deliberation and reasoning should be central in exemplar-applied moral education.

As a general note regarding the impacts of higher-level factors, people should invest efforts to establish sociocultural atmospheres where people highly regard moral values and virtues. If conventions and norms in a society positively reinforce anti-moral conduct, such conventions and norms demonstrated via people's behaviors will become negative exemplars for students (Engelen et al., 2018; Niebuhr, 2013). Such a situation may limit the positive influences of individual moral exemplars at the local level. Moral educators should also employ moral exemplars from diverse backgrounds to facilitate contextually sensitive moral development among diverse students while preventing the growth of bias,

prejudice, and discrimination (Čehajić-Clancy & Bilewicz, 2021; Murry et al., 2024). Problems regarding potential biases in AI models trained with large-scale datasets and solutions for dealing with them that we have overviewed so far may support the abovementioned general suggestion.

Concluding Remarks

In this paper, I discussed why and how employing exemplars plays a fundamental role in moral education. To support my point, I reviewed recent advances in AI research that can provide insights into understanding human cognition and learning processes at the behavioral and neural levels. I focused on studies reporting that exemplar-based learning, not rule-based learning, effectively promotes the development of capacities in diverse domains with flexibility. Based on these, I proposed that moral educators use exemplars and role models to facilitate moral development, particularly the development of complicated moral functioning, such as *phronesis*. In addition to the benefits, I also overviewed and discussed the potential issues involving using exemplars based on the Ecological Systems Theory, which explains the impacts of larger systems surrounding individuals. The review of recent AI research, particularly research on training biased LLMs and employing the CoT, suggests that moral educators utilize educational methods accompanied by autonomous, deliberative, and reasoning processes. Such will enable maximizing the positive outcomes of exemplar-applied moral education while minimizing its negative consequences, such as negative and biased modeling via influences from the meso- and macrosystems, as suggested by AI research.

Although I could discuss several practical points to improve the effectiveness of exemplar-applied moral education with evidence from AI research, I do not assume this

paper will perfectly address all concerns among moral educators. First, several researchers raised concerns that cognitive and learning mechanisms found among AI models might not sufficiently explain those among humans (Demszky et al., 2023). Although I agree that AI research will inform psychological and neuroscientific studies to understand human cognition and learning, given the potential limitation, we should be careful when we apply findings in computer science to research on moral education. Second, since the purpose of this paper was to overview AI research relevant to moral education and briefly examine their implications on moral education, I could not delve into further details about further details related to moral philosophy and psychology and concrete educational practice. Future works should address such limitations to provide moral educators with more insights into the relevant theories and educational methods.

Finally, I did not address whether cognitive processes in AIs and humans are fundamentally identical to each other. Although I argued that my point about the importance of using exemplars can still be supported at the practical level despite the differences, such might potentially weaken the foundation of my point. Researchers will need to pay attention to future studies in AI, particularly those examining more complicate cognitive capacities in AI, such as emotional processes (e.g., Hadi et al., 2023), to be able to address this concern better. Moreover, although it is out of the scope of the present study, we may also refer to philosophical critiques to the Chinese Room argument (see Cole, 2020). They can potentially provide additional support for my point based on assumptions that AI is possibly capable of human cognitive processes, such as intentionality.

Financial statement

I confirm that I have no relevant financial interests in relation to this publication.

Declaration of Competing Interest

None.

References

Abid, A., Farooqi, M., & Zou, J. (2021). Persistent Anti-Muslim Bias in Large Language Models. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 298–306. https://doi.org/10.1145/3461702.3462624

Aharoni, E., Fernandes, S., Brady, D. J., Alexander, C., Queen, K., Rando, J., ... & Crespo, V. (2024). Attributions toward Artificial Agents in a modified Moral Turing Test. https://doi.org/10.31234/osf.io/4gxtk

Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: The 'other-praising'emotions of elevation, gratitude, and admiration. *The Journal of Positive Psychology*, *4*(2), 105–127.

Alves, C. L., Toutain, T. G. L. D. O., Porto, J. A. M., Aguiar, P. M. D. C., De Sena, E. P., Rodrigues, F. A., Pineda, A. M., & Thielemann, C. (2023). Analysis of functional connectivity using machine learning and deep learning in different data modalities from individuals with schizophrenia. *Journal of Neural Engineering*, *20*(5), 056025. https://doi.org/10.1088/1741-2552/acf734

Athanassoulis, N. (2022). The Phronimos as a moral exemplar: Two internal objections and a proposed solution. *The Journal of Value Inquiry*. https://doi.org/10.1007/s10790-021-09872-4

Athanassoulis, N., & Han, H. (2023). Role Modeling is Beneficial in Moral Character Education: A Commentary on Carr (2023). *Philosophical Inquiry in Education*, *30*(3), 240–243.

Bandura, A., & McDonald, F. J. (1963). Influence of social reinforcement and the behavior of models in shaping children's moral judgments. *Journal of Abnormal and Social Psychology*, *67*(3), 274–281.

Bandura, A., Ross, D., & Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *The Journal of Abnormal and Social Psychology*, *67*(6), 601–607. https://doi.org/10.1037/h0045550

Binz, M., Dasgupta, I., Jagadish, A. K., Botvinick, M., Wang, J. X., & Schulz, E. (2023). Meta-Learned Models of Cognition. *Behavioral and Brain Sciences*, 1–38. https://doi.org/10.1017/S0140525X23003266

Blatt, M. M., & Kohlberg, L. (1975). The Effects of Classroom Moral Discussion upon Children's Level of Moral Judgment. *Journal of Moral Education*, *4*(2), 129–161. https://doi.org/10.1080/0305724750040207

Bronfenbrenner, U. (1993). Ecological Models of Human Development. In M. Gauvain & M. Cole (Eds.), *Readings on the Development of Children, 2nd Ed.* (pp. 37–43). Freeman.

Campolo, A., & Schwerzmann, K. (2023). From rules to examples: Machine learning's type of authority. *Big Data & Society*, *10*(2), 20539517231188725. https://doi.org/10.1177/20539517231188725

Carr, D. (2023). The Hazards of Role Modelling for the Education of Moral and/or Virtuous Character. *Philosophical Inquiry in Education*, *30*(1), 68–79.

Čehajić-Clancy, S., & Bilewicz, M. (2020). Appealing to Moral Exemplars: Shared Perception of Morality as an Essential Ingredient of Intergroup Reconciliation. *Social Issues and Policy Review*, *14*(1), 217–243. https://doi.org/10.1111/sipr.12067

Čehajić-Clancy, S., & Bilewicz, M. (2021). Moral-Exemplar Intervention: A New Paradigm for

Conflict Resolution and Intergroup Reconciliation. *Current Directions in Psychological Science*, *30*(4), 335–342. https://doi.org/10.1177/09637214211013001

Chalmers, D. J. (2023). *Could a Large Language Model be Conscious?* (arXiv:2303.07103). arXiv. http://arxiv.org/abs/2303.07103

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2023). *A Survey on Evaluation of Large Language Models* (arXiv:2307.03109). arXiv. http://arxiv.org/abs/2307.03109

Chen, J. H., Dhaliwal, G., & Yang, D. (2022). Decoding Artificial Intelligence to Achieve Diagnostic Excellence: Learning From Experts, Examples, and Experience. *JAMA*, *328*(8), 709. https://doi.org/10.1001/jama.2022.13735

Cole, D. (2020). The Chinese Room argument. In E. N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Stanford University. https://plato.stanford.edu/entries/chinese-room

Cowan, R. J., & Swearer, S. M. (2004). School–Community Partnerships. In *Encyclopedia of Applied Psychology* (pp. 309–317). Elsevier. https://doi.org/10.1016/B0-12-657410-3/00787-X

Damon, W., & Colby, A. (2013). Why a true account of human development requires exemplar research. In M. K. Matsuba, P. E. King, & K. C. Bronk (Eds.), *Exemplar methods and research: Quantitative and qualitative strategies for investigation. New Directions in Child and Adolescent Development* (pp. 13–26). Jossey-Bass.

Darnell, C., Fowers, B. J., & Kristjánsson, K. (2022). A multifunction approach to assessing

Aristotelian phronesis (practical wisdom). *Personality and Individual Differences*, *196*, 111684. https://doi.org/10.1016/j.paid.2022.111684

Darnell, C., Gulliford, L., Kristjánsson, K., & Paris, P. (2019). Phronesis and the Knowledge-Action Gap in Moral Psychology and Moral Education: A New Synthesis? *Human Development*, *62*(3), 101–129. https://doi.org/10.1159/000496136

De Caro, M., Marraffa, M., & Vaccarezza, M. S. (2021). The priority of phronesis: How to rescue virtue theory from its crisis. In M. De Caro & M. S. Vaccarezza (Eds.), *Practical Wisdom: Philosophical and Psychological Perspectives* (pp. 29–51). Routledge.

De Caro, M., Vaccarezza, M. S., & Niccoli, A. (2018). Phronesis as Ethical Expertise: Naturalism of Second Nature and the Unity of Virtue. *The Journal of Value Inquiry*, *52*(3), 287–305. https://doi.org/10.1007/s10790-018-9654-9

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., JonesMitchell, N., Ong, D. C., Dweck, C. S., Gross, J. J., & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, *2*(11), 688–701. https://doi.org/10.1038/s44159-023-00241-5

Dhar, V. (2023). *The Paradigm Shifts in Artificial Intelligence* (arXiv:2308.02558). arXiv. http://arxiv.org/abs/2308.02558

Dougherty, M. R., & Robey, A. (2018). Neuroscience and education: A bridge astray?. *Current Directions in Psychological Science, 27*(6), 401-406. https://doi.org/10.1177/0963721418794495

Engelen, B., Thomas, A., Archer, A., & Van De Ven, N. (2018). Exemplars and nudges: Combining two strategies for moral education. *Journal of Moral Education*, *47*(3),

346–365. https://doi.org/10.1080/03057240.2017.1396966

FeldmanHall, O., & Lamba, A. (2023). Learning to weigh competing moral motivations. In M. K. Berg & E. C. Chang (Eds.), *Motivation and morality: A multidisciplinary approach.* (pp. 157–184). American Psychological Association. https://doi.org/10.1037/0000342-007

Forbes, V. C. B., Dickinson, K. J. M., & Hulbe, C. L. (2022). Applying a social-ecological systems lens to patterns of policy, operational change, and gender participation in a large Aotearoa New Zealand organisation. *Journal of the Royal Society of New Zealand*, *52*(5), 539–568. https://doi.org/10.1080/03036758.2021.2012489

Frank, M. C. (2023). *Large language models as models of human cognition* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/wxt69

Freeman, D., Aquino, K., & McFerran, B. (2009). Overcoming beneficiary race as an impediment to charitable donations: Social dominance orientation, the experience of moral elevation, and donation behavior. *Personality and Social Psychology Bulletin*, *35*(1), 72–84.

Fridland, E., & Moore, R. (2015). Imitation reconsidered. *Philosophical Psychology, 28*(6), 856-880. https://doi.org/10.1080/09515089.2014.942896

Ganguli, D., Askell, A., Schiefer, N., Liao, T. I., Lukošiūtė, K., Chen, A., Goldie, A., Mirhoseini, A., Olsson, C., Hernandez, D., Drain, D., Li, D., Tran-Johnson, E., Perez, E., Kernion, J., Kerr, J., Mueller, J., Landau, J., Ndousse, K., … Kaplan, J. (2023). *The Capacity for Moral Self-Correction in Large Language Models* (arXiv:2302.07459). arXiv. http://arxiv.org/abs/2302.07459

Giannakidou, A., & Mari, A. (2024). The Human and the Mechanical: logos, truthfulness, and

ChatGPT. arXiv. https://arxiv.org/abs/2402.01267

Hacker-Wright, J. (2023). The Practical Unity of Practical Wisdom. *Topoi*. https://doi.org/10.1007/s11245-023-09999-y

Hadi, M. U., Tashi, Qasem al, Qureshi, R., Shah, A., Muneer, Amgad, Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., & Mirjalili, S. (2023). Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects. https://doi.org/10.36227/techrxiv.23589741.v4

Haidt, J. (2000). The positive emotion of elevation. *Prevention and Treatment*, *3*(3), 1–5.

Hamilton, B. (2019). Navigating Moral Struggle: Toward a Social Model of Exemplarity. *Journal of Religious Ethics*, *47*(3), 566–582. https://doi.org/10.1111/jore.12276

Han, H. (2015). Virtue ethics, positive psychology, and a new model of science and engineering ethics education. *Science and Engineering Ethics*, *21*(2), 441–460. https://doi.org/10.1007/s11948-014-9539-7

Han, H. (2023a). Considering the Purposes of Moral Education with Evidence in Neuroscience: Emphasis on Habituation of Virtues and Cultivation of Phronesis. *Ethical Theory and Moral Practice*. https://doi.org/10.1007/s10677-023-10369-1

Han, H. (2023b). Potential benefits of employing large language models in research in moral education and development. *Journal of Moral Education*, 1–16. https://doi.org/10.1080/03057240.2023.2250570

Han, H. (2024a). Examining Phronesis Models with Evidence from the Neuroscience of Morality Focusing on Brain Networks. *Topoi*. https://doi.org/10.1007/s11245-023-10001-y

Han, H. (2024b). Examining the Network Structure among Moral Functioning Components

with Network Analysis. *Personality and Individual Differences*, *217*, 112435. https://doi.org/10.1016/j.paid.2023.112435

Han, H., & Dawson, K. J. (2023). Relatable and attainable moral exemplars as sources for moral elevation and pleasantness. *Journal of Moral Education*. https://doi.org/10.1080/03057240.2023.2173158

Han, H., & Graham, M. (2023). *Considerations for Effective Use of Moral Exemplars in Education: Based on the Self-Determination Theory and Data Syntheses* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/n6mkp

Han, H., Kim, J., Jeong, C., & Cohen, G. L. (2017). Attainable and Relevant Moral Exemplars Are More Effective than Extraordinary Exemplars in Promoting Voluntary Service Engagement. *Frontiers in Psychology*, *8*, 283. https://doi.org/10.3389/fpsyg.2017.00283

Han, H., Lee, K., & Soylu, F. (2020). Applying the Deep Learning Method for Simulating Outcomes of Educational Interventions. *SN Computer Science*, *1*(2), 70. https://doi.org/10.1007/s42979-020-0075-z

Han, H., Soylu, F., & Anchan, D. M. (2019). Connecting Levels of Analysis in Educational Neuroscience: A Review of Multi-level Structure of Educational Neuroscience with Concrete Examples. *Trends in Neuroscience and Education*, 100113. https://doi.org/10.1016/j.tine.2019.100113

Han, H., Workman, C. I., May, J., Scholtens, P., Dawson, K. J., Glenn, A. L., & Meindl, P. (2022). Which moral exemplars inspire prosociality? *Philosophical Psychology*, *35*(7), 943–970. https://doi.org/10.1080/09515089.2022.2035343

Henderson, E. (2023). Entangled *phronesis* and the four causes of emulation:

Developmental insights into role modelling. *Theory and Research in Education*, *21*(3), 264–283. https://doi.org/10.1177/14778785231203104

Henderson, E. (2024a). The educational salience of emulation as a moral virtue. *Journal of Moral Education, 53*(1), 73–88. https://doi.org/10.1080/03057240.2022.2130882

Henderson, E. (2024b). Special issue on exemplars and emulation in moral education: Guest editorial. *Journal of Moral Education, 53*(1), 1-13. https://doi.org/10.1080/03057240.2023.2291213

Ho, N., Schmid, L., & Yun, S.-Y. (2023). *Large Language Models Are Reasoning Teachers* (arXiv:2212.10071). arXiv. http://arxiv.org/abs/2212.10071

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Ke, L., Tong, S., Cheng, P., & Peng, K. (2024). *Exploring the Frontiers of LLMs in Psychological Applications: A Comprehensive Review* (arXiv:2401.01519). arXiv. http://arxiv.org/abs/2401.01519

Koppe, G., Meyer-Lindenberg, A., & Durstewitz, D. (2021). Deep learning for small and big data in psychiatry. *Neuropsychopharmacology*, *46*(1), 176–190. https://doi.org/10.1038/s41386-020-0767-z

Kosinski, M. (2023). *Theory of Mind May Have Spontaneously Emerged in Large Language Models* (arXiv:2302.02083). arXiv. http://arxiv.org/abs/2302.02083

Kristjánsson, K. (2006). Emulation and the use of role models in moral education. *Journal of Moral Education*, *35*(1), 37–49.

Kristjánsson, K. (2014). Phronesis and moral education: Treading beyond the truisms. *Theory and Research in Education*. https://doi.org/10.1177/1477878514530244

Kristjánsson, K. (2017). Emotions targeting moral exemplarity: Making sense of the logical geography of admiration, emulation and elevation. *Theory and Research in Education*, *15*(1), 20–37. https://doi.org/10.1177/1477878517695679

Kristjánsson, K., & Fowers, B. (2022). Phronesis as moral decathlon: Contesting the redundancy thesis about phronesis. *Philosophical Psychology*, 1–20. https://doi.org/10.1080/09515089.2022.2055537

Kristjánsson, K., Fowers, B., Darnell, C., & Pollard, D. (2021). Phronesis (Practical Wisdom) as a Type of Contextual Integrative Thinking. *Review of General Psychology*, *25*(3), 239–257. https://doi.org/10.1177/10892680211023063

Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M., & Kanai, R. (2022). Meta-learning, social cognition and consciousness in brains and machines. *Neural Networks*, *145*, 80–89. https://doi.org/10.1016/j.neunet.2021.10.004

LeCun, Y., Bengio, Y., Hinton, G., Goodfellow, I. J., & Courville, A. (2015). Deep Learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lengbeyer, L. (2022). Dismantling the Chinese Room with linguistic tools: a framework for elucidating concept-application disputes. *AI & SOCIETY*, 37, 1625-1643. https://doi.org/10.1007/s00146-021-01257-2

Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., Lyons, K., Gašević, D., & Chen, G. (2023). Can large language models write reflectively. *Computers and Education:*

Artificial Intelligence, *4*, 100140. https://doi.org/10.1016/j.caeai.2023.100140

Liévin, V., Hother, C. E., Motzfeldt, A. G., & Winther, O. (2023). *Can large language models reason about medical questions?* (arXiv:2207.08143). arXiv. http://arxiv.org/abs/2207.08143

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology*, *83*(4), 854–864. https://doi.org/10.1037/0022-3514.83.4.854

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, *73*, 91–103. https://doi.org/10.1037/0022-3514.73.1.91

Macpherson, T., Churchland, A., Sejnowski, T., DiCarlo, J., Kamitani, Y., Takahashi, H., & Hikida, T. (2021). Natural and Artificial Intelligence: A brief introduction to the interplay between AI and neuroscience research. *Neural Networks*, *144*, 603–613. https://doi.org/10.1016/j.neunet.2021.09.018

Mu, Y., Zhang, Q., Hu, M., Wang, W., Ding, M., Jin, J., Wang, B., Dai, J., Qiao, Y., & Luo, P. (2023). *EmbodiedGPT: Vision-Language Pre-Training via Embodied Chain of Thought* (arXiv:2305.15021). arXiv. http://arxiv.org/abs/2305.15021

Murry, V. M., Hanebutt, R., Han, H., Debreaux, M., & Nyanamba, J. (2024). *Culturally Sensitive and Contextually Adapted Exemplars of Character Development: Implications for Reimagining Frameworks*.

Naous, T., Ryan, M. J., Ritter, A., & Xu, W. (2023). *Having Beer after Prayer? Measuring Cultural Bias in Large Language Models* (arXiv:2305.14456). arXiv.

http://arxiv.org/abs/2305.14456

Nartey, P., Bahar, O. S., & Nabunya, P. (2023). A review of the cultural gender norms contributing to gender inequality in Ghana: An Ecological Systems perspective. *Journal of International Women's Studies*, *25*(7), 14.

Navigli, R., Conia, S., & Ross, B. (2023). Biases in Large Language Models: Origins, Inventory, and Discussion. *Journal of Data and Information Quality*, *15*(2), 1–21. https://doi.org/10.1145/3597307

Niebuhr, R. (2013). *Moral man and immoral society: A study in ethics and politics* (2. ed). Westminster John Knox Press.

Nolfi, S. (2023). *On the Unexpected Abilities of Large Language Models* (arXiv:2308.09720). arXiv. http://arxiv.org/abs/2308.09720

Penn, W. Y. (1990). Teaching Ethics -A Direct Approach. *Journal of Moral Education*, *19*(2), 124–138. https://doi.org/10.1080/0305724900190206

Piel, M. H., Geiger, J. M., Julien-Chinn, F. J., & Lietz, C. A. (2017). An ecological systems approach to understanding social support in foster family resilience. *Child & Family Social Work*, *22*(2), 1034–1043. https://doi.org/10.1111/cfs.12323

Prystawski, B., Thibodeau, P., Potts, C., & Goodman, N. D. (2023). *Psychologically-informed chain-of-thought prompts for metaphor understanding in large language models* (arXiv:2209.08141). arXiv. http://arxiv.org/abs/2209.08141

Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. A., & Botvinick, M. (2018). Machine theory of mind. *In Proceedings of the 35th International Conference on Machine Learning* (pp. 4218-4227). PMLR.

Rashid, M., Singh, H., & Goyal, V. (2020). The use of machine learning and deep learning

algorithms in functional magnetic resonance imaging—A systematic review. *Expert Systems*, *37*(6), e12644. https://doi.org/10.1111/exsy.12644

Redford, J. L., Mcpherson, R. H., Frankiewicz, R. G., & Gaa, J. (1995). Intuition and Moral Development. *Journal of Psychology*, *129*(1), 91–101.

Sanderse, W. (2012). The meaning of role modelling in moral and character education. *Journal of Moral Education*, *42*(1), 28–42. https://doi.org/10.1080/03057240.2012.690727

Sanderse, W. (2023). Adolescents' moral self-cultivation through emulation: Implications for modelling in moral education. *Journal of Moral Education*. https://doi.org/10.1080/03057240.2023.2236314

Saxe, A. M. (2015). *Deep Linear Neural Networks: A Theory of Learning in the Brain and Mind*. Stanford University.

Schnall, S., Roper, J., & Fessler, D. M. T. (2010). Elevation leads to altruistic behavior. *Psychological Science*, *21*, 315–320. https://doi.org/10.1177/0956797609359882

Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A., & Kersting, K. (2022). *Large Pre-trained Language Models Contain Human-like Biases of What is Right and Wrong to Do* (arXiv:2103.11790). arXiv. http://arxiv.org/abs/2103.11790

Schwitzgebel, E., Schwitzgebel, D., & Strasser, A. (2023). Creating a Large Language Model of a Philosopher. *Mind & Language*. https://doi.org/10.1111/mila.12466

Shaikh, O., Zhang, H., Held, W., Bernstein, M., & Yang, D. (2023). *On Second Thought, Let's Not Think Step by Step! Bias and Toxicity in Zero-Shot Reasoning* (arXiv:2212.08061). arXiv. http://arxiv.org/abs/2212.08061

Silvers, J. A., & Haidt, J. (2008). Moral elevation can induce nursing. *Emotion*, *8*(2), 291–295.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Scharli, N., Chowdhery, A., Mansfield, P., Arcas, B. A. y, Webster, D., … Natarajan, V. (2022). *Large Language Models Encode Clinical Knowledge* (arXiv:2212.13138). arXiv. http://arxiv.org/abs/2212.13138

Smith, R. H. (2000). Assimilative and contrastive emotional reactions to upward and downward social comparisons. In J. Suls & L. Wheeler (Eds.), *Handbook of Social Comparison: Theory and Research* (pp. 173–200). Kluwer Academic/Plenum Publishers.

Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., Kluska, A., Lewkowycz, A., Agarwal, A., Power, A., Ray, A., Warstadt, A., Kocurek, A. W., Safaya, A., Tazarv, A., … Wu, Z. (2023). *Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models* (arXiv:2206.04615). arXiv. http://arxiv.org/abs/2206.04615

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science*, *11*, 159–163. https://doi.org/10.1111/1467-8721.00191

Tian, J.-J., Dige, O., Emerson, D., & Khattak, F. (2023). Using Chain-of-Thought Prompting for Interpretable Recognition of Social Bias. *Socially Responsible Language Modelling Research*. https://openreview.net/forum?id=QyRganPqPz

Trott, S., Jones, C., Chang, T., Michaelov, J., & Bergen, B. (2023). Do Large Language Models Know What Humans Know? *Cognitive Science*, *47*(7), e13309. https://doi.org/10.1111/cogs.13309

Vaccarezza, M. S., Kristjánsson, K., & Croce, M. (2023). Phronesis (Practical Wisdom) as a Key to Moral Decision-Making: Comparing Two Models. *The Jubilee Centre for Character & Virtues Insight Series*.

Urman, A., & Makhortykh, M. (2023). The Silence of the LLMs: Cross-Lingual Analysis of Political Bias and False Information Prevalence in ChatGPT, Google Bard, and Bing Chat. https://doi.org/10.31219/osf.io/q9v8f

Vaccarezza, M. S., & Niccoli, A. (2019). The dark side of the exceptional: On moral exemplars, character education, and negative emotions. *Journal of Moral Education*, *48*(3), 332–345. https://doi.org/10.1080/03057240.2018.1534089

Vaccarezza, M. S., & Niccoli, A. (2022). Let the donkeys be donkeys: In defense of inspiring envy. In S. Protasi (Ed.), *The Moral Psychology of Envy* (pp. 111–127). Rowman & Littlefield.

Van Der Waa, J., Nieuwburg, E., Cremers, A., & Neerincx, M. (2021). Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence*, *291*, 103404. https://doi.org/10.1016/j.artint.2020.103404

Van Dijk, B., Kouwenhoven, T., Spruit, M. R., & van Duijn, M. J. (2023). Large language models: The need for nuance in current debates and a pragmatic perspective on understanding. arXiv. https://arxiv.org/abs/2310.19671

Vianello, M., Galliani, E. M., & Haidt, J. (2010). Elevation at work: The effects of leaders' moral excellence. *The Journal of Positive Psychology*, *5*(5), 390–411.

Wang, B., Li, G., & Li, Y. (2023). Enabling Conversational Interaction with Mobile UI using Large Language Models. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3544548.3580895

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models* (arXiv:2201.11903). arXiv. http://arxiv.org/abs/2201.11903

Williams, J., Fiore, S. M., & Jentsch, F. (2022). Supporting Artificial Social Intelligence With Theory of Mind. *Frontiers in Artificial Intelligence*, *5*, 750763. https://doi.org/10.3389/frai.2022.750763

Yamazaki, T., Yoshikawa, K., Kawamoto, T., Mizumoto, T., Ohagi, M., & Sato, T. (2023). Building a hospitable and reliable dialogue system for android robots: A scenario-based approach with large language models. *Advanced Robotics*, *37*(21), 1364–1381. https://doi.org/10.1080/01691864.2023.2244554

Zagzebski, L. (2013). Moral exemplars in theory and practice. *Theory and Research in Education*, *11*(2), 193–206. https://doi.org/10.1177/1477878513485177

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2023). *A Survey of Large Language Models* (arXiv:2303.18223). arXiv. http://arxiv.org/abs/2303.18223