

EXTENDED VIRTUES AND FUTURE ETHICS

XIUYUAN AN

SCHOOL OF PHILOSOPHY, FUDAN UNIVERSITY

ABSTRACT

The Extended Virtues (EV) theory challenges the traditional notion that an individual's character traits stem solely from their mental state and posits that external factors also play a big part in how we characterize virtues. In critiquing the philosopher Robert Howell's version of EV, this article advocates for EV by examining the essential constituents of virtues. It emphasizes the necessity of consistent moral decision-making and argues that virtues are better characterized through the active engagement of agents rather than being passively influenced by external conditions. The article also envisions future ethical scenarios where virtues are dynamically integrated with technology and social systems.

KEY WORDS

Extended mind, extended virtues, ethics, moral agent, moral decision-making.

METADATA

doi.org/10.4245/sponge.v12i1.24245

CONTACT

18340160004@fudan.edu.cn

INTRODUCTION

The traditionally accepted view of characters and virtues, known as mentalism, is that people's character traits depend entirely on their mental states (Annas, 2011). Some philosophers believe that nonmental features also play a role in forming characters and virtues. Despite their disagreements regarding the role of nonmental features in character and virtue formation, both would likely embrace skindividualism, the view that a person's virtue must involve a disposition that is wholly grounded by features inside the person's skin (Howell, 2016, p. 146). More specifically, skindividualism posits that virtues are solely constituted by mental dispositions, acknowledging that socio-cultural factors influence moral dispositions but maintaining that these virtues reside within the individual's mental states. In opposition to skindividualism, the Extended Virtues (EV) view holds that one's virtues do not supervene on what is inside a person's skin (Howell, 2016, p. 154), but can have an extended component such as the environment (see Olson, 2011; Olson & Witt, 2020; Skorburg, 2019). According to this view, virtues are not just internal states but also encompass the external contexts and situations that influence moral behaviour, thereby challenging the traditional

assumption that dispositions responsible for virtues must be entirely grounded in the internal features of the person.

The philosopher Robert J. Howell (2016) draws on findings from situationist psychology to argue that, if virtues are to exist, they must be extended. He suggests that these virtues are supported by systems external to ourselves, which challenges the notion of skindividualism and points to the possibility that virtues, like minds, extend beyond the individual (p. 148). This argument is inspired by the *Extended Mind* (EM) thesis put forward by Clark and Chalmers (1998), which proposes that the mind is not confined within the skin and skull but extends beyond the body into the environment. Howell contends that, while not all psychological states may be extended, the system encompassing virtues operates as a unified agent with components extending beyond the individual, thus extending their virtues.

This article argues that moral decision-making extends virtues. My argument differs from Howell's EV argument by emphasizing that, apart from external conditions that can be counted to extend virtues, the individual's participation in a specific action is compulsory in justifying their virtues. This highlights that the

agent's decision process includes motivation and reasoning, which play a crucial role in consistent moral actions. The article is structured as follows. The first section elaborates on Howell's EV thesis in detail and points out the shortcomings of his argument for EV. Section 2 discusses the conditions of virtues to characterize EV. Then, section 3 presents an alternative approach to EV that requires individuals to make consistent decisions in moral actions. The final section envisions some future ethical scenarios as positive examples of EV.

EXTENDED VIRTUES: AN EVALUATION

According to skindividualism, an individual's virtue is wholly grounded by features inside the person's skin (Howell, 2016, p. 146). This view implies that people with the same internal configurations are supposed to have the same virtues regardless of the external situation they find themselves in. However, doubts have been raised about the robustness of certain character traits. One notable example is the series of obedience experiments conducted by the psychologist Stanley Milgram in the 1960s. In these experiments, participants were instructed to administer increasingly severe electric shocks to another person (who was actually a confederate acting like a regular participant) whenever that person made a mistake. Despite the apparent suffering of the person being shocked, a significant number of participants followed the experimenter's commands and administered shocks at extreme levels. Milgram's study revealed that many participants were willing to go all the way with the shocks not because they had a predisposition for cruelty, but because they were obeying the authority figure in the experiment (Milgram, 1974).

Situationism — that it is the situation leads ordinary people to do terrible things even if their internal configurations are alike — provides a more plausible explanation. The same logic applies to EV: one's virtues do not supervene upon what is inside a person's skin but can have an extended component (e.g., the environment) that accounts for differences in behaviours and virtues. Howell draws on situationist psychology to support the claim that an individual's character

traits are not confined to their internal states. He argues that the social environment, including other people and surrounding technologies, plays a crucial role in the formation of character. According to Howell, these external factors are essential in sustaining the virtues that shape our characters, noting that these traits are supported by a variety of external factors and generally persist even when we are alone. Slaby and Gallagher (2015), Sneddon (2011), Cash (2013), Carter et al. (2016), and Alfano and Skorburg (2017) offer similar views, highlighting the importance of external factors in shaping virtues. Howell (2016) further argues that if the external grounds of a virtuous disposition are stable enough, "there is less need for internal sources of moral power." (p. 160) To illustrate how external factors influence virtues, Howell (2016) presents the analogy of a solar-powered traffic light, which helps clarify the impact of such external supports on virtuous behavior:

"There are many ways to build such a light, not all of them equally good. A particularly bad design would be a solar traffic light that was only powered by the sun's rays during the equinox. Even if it worked well in that rare perpendicular light, it would not be a particularly good traffic light." (p. 159)

According to Howell (2016), a solar-powered traffic light is a "bad" light because it only functions on sunny days. Still, it can be a "good" light if it is installed in a place where the sun never sets (p. 160). Here, a solar-powered battery is the internal configuration of this light, whereas the sun and other weather statuses are the external conditions. We can say that these two components together decide the traffic light's "moral nature." Howell's traffic light analogy illustrates the need to understand external factors in evaluating virtues, and the degree of virtues to be extended is a function of circumstances. In a possible world where the sun never sets, a solar-powered traffic light can function well without an internal battery. Similarly, in a place with intensive surveillance, all "bad" behaviours are

stopped before they can occur; we can be virtuous *without* internal moral motivation. Howell's ultimate goal is to set standards for evaluating virtues on the ways of being a person. This is in response to the potential critique of his focus on conscious states, with the acknowledgement that a person is not solely defined by conscious states.

While I find the idea of EV plausible, I am not fully committed to Howell's argument. Howell primarily elucidates observable and overt behaviours (e.g., virtuous actions one takes), but this viewpoint has two shortcomings. On the one hand, external environmental factors might be unstable. I agree that we need to acknowledge the significance of external influences in shaping virtues. But if virtues depend on environmental factors that can be easily changed, then the repercussion is that virtues are also unstable. This necessitates proving the effectiveness of external factors to form the virtues. When assessing the virtue of a solar-powered traffic light, we need to consider the sunlight and other contributing factors like technology, weather, and the light's structural robustness. All of these factors must collectively function to form a unified entity, or else the virtue of an individual would not be sustained. Likewise, an agent's moral choices will not reliably come from one's motivation or reason but are contingent on changeable external sources, which weakens Howell's claim.

On the other hand, a key issue with Howell's argument is that it presupposes the relevance of the external environment in constituting virtues. His position holds only if the external conditions are pertinent to the virtue in question. This creates a potential problem: if external factors are irrelevant, then they should not influence our judgment of virtues. For instance, when evaluating the functioning of a solar-powered traffic light, the relevant external factor is the intensity of daylight, not the presence of wind or rain. According to Howell, only factors like daylight intensity matter for the light's operation, while irrelevant conditions do not contribute to its virtue. This raises the question of whether, in the case of human virtues, we can always be sure that the external factors we rely on are truly relevant to the disposition in question.

Howell's focus on observable and overt behaviours of virtues leaves crucial questions about dispositions and motivations that drive these virtues unanswered. Virtues are indeed influenced by external factors. However, if virtues are heavily influenced by external factors regardless of dispositions and motivations, potential conflicts may arise. Often, we do not know what surrounding environment influences our dispositions; it can be due to an occasional hunger or a sleepless night, or there is no reason at all. These influences could lead to the consequence that what is considered virtuous in one context could be deemed otherwise in another. A brave soldier may go forward fearlessly on the battlefield, yet they might be startled when dealing with cockroaches in their home. This indicates that certain contextualized dispositions, though noteworthy, may not entirely encapsulate the reasons or motivations behind actions across diverse circumstances.

Therefore, while I acknowledge Howell's insights into the external influences on virtues, I argue that a comprehensive understanding requires a more nuanced consideration of the interaction between internal dispositions and external factors. In this way, virtues can maintain a degree of stability and reliability, ensuring they serve as a consistent guide for moral actions. The subsequent section will continue this analysis of the nature of virtues.

WHAT ARE EXTENDED VIRTUES

To characterize the virtues of a traffic light in the extended sense, one may ask: what makes a virtue? But this question is equivocal. It could concern the functional role to which we attribute praise or blame, or pertain to what constitutes the essential nature of virtues. This article focuses on the latter. To understand what extended virtues are, I suggest considering virtues in terms of two aspects: their functional role and their essential constituents.

When considering the functional role of virtues, we tend to focus on what virtues do and how they operate. For example, a generous person gives to those in need, a courageous person stands firm in the face of danger, and a just person treats others fairly. In this sense, virtues are defined by

the actions and behaviours they produce. However, merely examining the functional role is insufficient. Because virtues cannot always be instantiated due to an individual's inability to carry out the virtuous action or the circumstances prevent such instantiations of virtues. A virtue, therefore, is not only a matter of what one does but also involves the conditions under which one acts.

Regarding the essential constituents of virtues, we must also include both internal states (e.g., intentions and desires) and external conditions (e.g., social and environmental factors). For example, assessing the virtues of athletes involves observing their actions on the playing field, which we believe reflect their internal states. Such an assessment, however, also involves their countless hours of training, unwavering commitment, and the courage they demonstrate in pushing beyond their limits during competition. Essentially, an athlete's virtue is found in their relentless pursuit of a goal. The display of qualities in any single action embodies the effort and dedication applied across various situations.

Additionally, when considering the effect of the socio-cultural environment on the agent, we should be selective rather than accepting all influences indiscriminately. It seems to me that agents' identities and abilities play a more important role than understanding their environment and accepting social norms in initiating virtuous actions. People usually expect musicians to display calmness on stage, which is often regarded as a mark of professionalism. However, despite their talent, some musicians still experience extreme nervousness before performing. Here, being calm is a social expectation that acts as a normative virtue, but it should be secondary to characters more directly related to musicians' identities and abilities. They might resort to additional preparation or even neuro-doping to manage this anxiety, but these measures are unrelated to their musical skills. Consequently, it is possible to accept that musicians can be nervous but still deliver great performances.¹

Nevertheless, virtues cannot be fully understood without understanding agents' actions

in specific scenarios. As with any process whose full causal story is unclear, changing the reasons that motivate an individual's actions can significantly reshape our judgment of the relationship between the individual and their moral character traits. Consider the following spilled coffee case. In this scenario, drinking coffee enhances productivity, and spilled coffee spoils the work that is done. Additionally, spilling coffee will stimulate this person to get extra productivity that compensates for the spoiled work. Without complete knowledge, there is a risk of making incorrect judgements about the virtue. After all, spilled coffee is a mistake that leads one to conclude that the person was careless. The negative consequence can be saved if the person continues to perform carefully and efficiently by drinking more coffee. In the latter scenario, the person may exhibit meticulousness with additional virtues, such as persistence and diligence.

The examination above shows that how we characterize virtues depends on virtuous actions that are described conditionally on both internal and external factors. What matters most, I have argued, is the active participation of agents. The EV framework defines virtues in terms of ongoing interactions with the environment. Accordingly, agents can adjust their actions based on feedback and continuously update their understanding of themselves and the world. Understanding the situation an agent is in also enables us to adapt our moral assessments and expectations of the agent. The spilled coffee case suggests that assessing the agent's virtues is not limited to a set of generalized cases but rather grounding their actions in every specific context. Virtues serve as explanations for individuals' behaviours, provided that their virtues are justified by their reasons and portrayed by a series of choices.

MAKING CONSISTENT DECISIONS IN MORAL ACTIONS

By presenting an alternative approach to EV, I argue that virtues should be understood as dispositions to bring about consistent moral decision processes. In emphasizing the moral decision-making process, this perspective highlights the importance of the active

engagement of moral agents with their environment. Virtues are not just passive characters but are actively developed through the choices individuals make. In what follows, I propose a view of EV that differs slightly from Howell's. This approach emphasizes that an agent's participation — through attention, effort and focus — is fundamental to making actions virtuous.

Suppose you received some love letters. Would you prefer ones crafted by human intelligence or ones generated by artificial intelligence? Most people would likely choose the human-crafted letter. However, if the choice were between a poorly written human-crafted three-line letter and a beautifully written 300-line AI letter, the decision might not be as clear-cut. In such a case, people might not prioritize whether the letter was human or AI-generated, especially if the latter successfully communicated genuine emotion. More importantly, if AI was instructed by a real human to craft such a love letter, and this human also edited the love letter to ensure it was suitable for the recipient, then it might be easier for the recipient to accept because the human's love is included, even with some help of AI! We, at least somebody, would likely choose AI-crafted letters rather than human-crafted ones just because the quality of the former is better. Generally, people appreciate well-written and emotionally resonant content, regardless of its source. Perhaps people value not only creativity, fluency, and emotional impact that can be attributed to any individual, but also the intention and resulting quality of the agent. It is acceptable to use AI tools to craft love letters, as long as the individual takes part in its production.

According to the above criticism of Howell's proposal about understanding agents' actions, it is the intrinsic value of personal effort and intention, rather than external factors, that support these virtues. Even if a letter is human-crafted, it will not be considered a good love letter without authenticity and personal touch. While it is important to consider virtues as a product of personhood shaped by an individual's internal states and external environments (i.e., an inevitable result), it is also crucial to examine the *moral decision-making* processes (i.e., a dynamic

integration of these effects) that form an individual's personhood.

Moral decision-making describes the process where individuals consistently choose actions aligned with moral values. It refers to the cognitive states that people experience when carrying out morally significant actions. My alternative argument for EV is motivated by the contention that due to human limitations and various circumstances, people may not always fully express their personhood, yet they can still make decisions before each action and find other supplementary tools to achieve their goals. This approach provides a stable foundation for characterizing virtues. The active participation of the agent in making moral decisions ensures that virtues are not merely responses to external conditions but involve consistent reasons.

My argument for the active engagement of virtuous actions is that we must have valid and necessary reasons for attributing virtuous extensions to ourselves rather than to someone else. What makes an action virtuous is not about the behaviour itself, nor the internal cognitive configurations or external influences that shape an individual. Rather, to possess a virtue, one must come to believe in the entire decision-making process, based on reasons that rationalize and justify it.² Without such reasoning, the result is that behaviours with little or no moral significance could be mistaken for virtuous actions. To better understand how virtues endure and adapt across different contexts, let us consider some related scenarios that illustrate how virtues interact with the environment.

The first case concerns the virtues of a Time Lord, specifically the Doctor, a millennia-old humanoid alien from the planet Gallifrey in the BBC science fiction television program *Doctor Who* (Davies, 2005). The Doctor is known for their ability to regenerate, a process that allows them to transform in both appearance and cellular structure after sustaining fatal injuries. Throughout their travels across time and space, the Doctor consistently performs benevolent acts, although sometimes introducing danger as well. Despite the risks involved, the Doctor is universally regarded as kind by human standards. In this case, the evaluative standards for kindness

remain stable, even though the Doctor undergoes radical biological changes during each regeneration. The Doctor's extended lifespan and robust vitality highlight how virtues can be influenced by different temporal and spatial contexts, complicating the evaluation of moral decisions. However, it is difficult to assess virtues based solely on biological properties, as we cannot be sure that functional or dispositional properties change alongside the Doctor's physical transformation. The Doctor's experience suggests that virtues are not merely internal dispositions or external factors. Instead, they involve the active participation of individuals in their circumstances. Virtue, therefore, is not a static trait but a dynamic process shaped by continuous engagement with the environment.

For comparison, consider a human surgeon who must decide whether to allocate limited resources to a patient. The virtue of compassion, in this case, is not merely a disposition to feel sympathy for the patient or a set of external factors that support the surgeon's actions. Rather, it involves the surgeon actively engaging with the situation, considering the patient's needs, the availability of resources, and the broader social context. The surgeon's virtue, therefore, is expressed through their decision-making process. What these two cases have in common is that if some reasons for decisions or actions remain unchanged under violations of circumstances, then it is enough to prove that this is the virtue of this person.

This view of virtues as dynamic processes emphasizes the importance of decision-making in moral actions. It requires individuals to actively engage with their environment, consider relevant factors, and make decisions that reflect their moral commitments. The EV thesis is not merely about extending the boundaries of virtues to include external factors. It is also about recognizing the active role of individuals in shaping their virtues through their decisions and actions. The importance of agency and responsibility in moral actions is highly valued, suggesting that virtues are not just about having the right internal dispositions or being in the right external circumstances, but also about making consistent decisions that align with moral values.

ENVISIONING FUTURE ETHICAL SCENARIOS

The above discussion about the nature of virtues reveals an interesting possibility for us to envision future ethical scenarios. In this section, I explore two potential future ethical scenarios that embody the principles of EV. These examples demonstrate how virtues can be extended and actively integrated into various aspects of life, including technologies, social systems, and environmental practices.

One possible future ethical scenario is the integration of AI and technology into moral decision-making processes. Imagine a future where AI tools are designed to help students with their studies. You are a philosophy student who wonders about the extended mind topic. You choose to use AI tools to formulate your ideas and construct arguments. AI technologies can provide information, offer suggestions, and help you to make more informed and virtuous decisions. Will this compromise your academic integrity? No. The use of AI in academia depends on its role in contributing to the student's studies. If AI tools are designed as tutors, it means they guide students to think critically rather than providing ready-made answers. Such functions are just like meeting with professors during their office hours. These uses of AI seem to fit within current academic rules and align with academic integrity. Therefore, whether the student using it has virtue depends on how and to what extent students use it.

Another example is the development of social systems that promote and support virtuous behaviours. Educational programs, community initiatives, and policies that encourage cooperation, empathy, and justice can extend virtues beyond individual actions to collective practices. We may further distinguish between contextualized and generalized virtues. Generalized virtues such as kindness, justice, perseverance, and courage coexist with contextually derived virtues such as kindness to enemies, distributive justice, and courage to face one's true self. Contextualized virtues hold intrinsic value and are not subordinate to generalized virtues. Below is an example to illustrate.

Suppose one uses AI tools to analyze

family members' preferences and automatically order gifts for them on birthdays. Naturally, one might assume that a person who genuinely cares about their family would thoughtfully handpick personalized gifts for them. However, the virtue of caring does not lie in *handpicking* by any particular person, but in understanding the preferences of the family and picking out what they like most. AI tools, if used appropriately, help individuals gain a better understanding of their preferences, rendering them an extension of virtues.³ Given our motivation to choose gifts for the people we love, the underlying reasons are inherently personal. Using AI tools to select gifts does not diminish their value compared to handpicking them. In fact, it can be viewed as an extension of the general virtue of caring for our families, contextualized within the modern capabilities of technology.

For various virtues in people's lives, there are different ways to contextualize a general virtue. We expect a person's character traits to reflect enduring intentions and consistent efforts. Utilizing advanced tools can enhance the stability and reliability of our decision-making processes. In other words, using tools is a means of engaging in virtuous actions, contributing to potential moral development. As Nancy Snow (2019) emphasizes, we need to work out new conceptions of virtues that adequately address the ethical demands of novel situations and contexts created by "advances in technology and science"

(p. 188). Thus, exercising our virtues with the aid of new technologies influences our understanding and evaluation of these virtues.

CONCLUSION

Virtues as an individual's consistent moral dispositions can be understood in two ways. The traditional view asserts that virtues are rooted solely in internal cognitive states, including functional, dispositional properties, and biological attributes. In contrast, the perspective of Howell's EV renews our understanding of virtues by incorporating external factors. Virtues are thus constituted by a complex interplay between an individual's internal states and external environments.

This article argues for a revised version of EV, stating that an individual's virtue is better characterized by including the motives and reasoning behind one's moral decisions. The active participation of agents — through their decision processes — is the essential constituent of virtue. Recognizing the decision-making process behind moral actions allows for a more nuanced consideration of the interaction between internal dispositions and external dynamics. With the help of future techniques as virtuous extensions, the EV view also supports moral development. Those techniques enable individuals to become fully realized moral agents despite their limitations, further promoting the agent's autonomy.

ENDNOTES

1. For more discussion, see Pugh & Pugh (2021).
2. Similar arguments regarding the role of reasons in forming beliefs and intentions can be found in Arpaly (2006) and Wedgwood (2006), who discuss how the normative force of reasons shapes our actions and beliefs, making them virtuous. What distinguishes virtuous actions is not merely the behavior itself, but the rational deliberation and reasons that justify and guide the action, ensuring it aligns with virtue.
3. Controversies about whether this suffices to render the use of such AI tools virtuous see Hernández Orallo & Vold (2019) and Andorno & Lavassa (2023).

ACKNOWLEDGEMENT

This paper began as a Neuroethics term paper in Fall 2022. I am deeply indebted to Jie Yin for her teaching, writing, and encouragement. Many thanks to Rachel Katz and Matthew McLaughlin,

organizers of the 2024 IHPST Graduate Workshop at the University of Toronto, for their support, and to my commentator, Jessie Hall, for her insightful suggestions and comments. I am also very grateful to Benjamin Winokur, Tomasz Wysocki, and participants of the Phil & Sci Reading Group and Philosophy of AI Summer School in Paris for their helpful feedback on earlier drafts. My heartfelt appreciation to Zili Dong for his invaluable assistance in refining this paper.

BIBLIOGRAPHY

- Alfano, M., & Skorburg, J. A. (2017). The Embedded and Extended Character Hypotheses. In J. Kiverstein (Ed.), *The Routledge Handbook of Philosophy of the Social Mind* (pp. 465–478). Routledge.
- Andorno, R. & Lavazza, A. (2023). How to deal with mind-reading technologies. *Frontiers in Psychology*, 14, 1290478.
- Annas, J. (2011). *Intelligent Virtue*. Oxford University Press.
- Arpaly, N. (2006). *Merit, meaning, and human bondage: An essay on free will*. Princeton University Press.
- Carter, J. A., Gordon, E. C., & Palermos, S. O. (2016). Extended emotion. *Philosophical Psychology*, 29(2), 198–217.
- Cash, M. (2013). Cognition without borders: “Third wave” socially distributed cognition and relational autonomy. *Cognitive Systems Research*, 61–71.
- Chalmers, D. (2019). Extended Cognition and Extended Consciousness. In M. Colombo, E. Irvine, & M. Stapleton (Eds.), *Andy Clark and his Critics*. Wiley-Blackwell.
- Clark, A., & Chalmers, D. J. (1998). The Extended Mind. *Analysis*, 58(1), 7–19.
- Davies, R. T. (Executive Producer). (2005–2010). *Doctor Who* [TV series]. BBC.
- Hernández-Orallo, J., & Vold, K. (2019). AI extenders: The ethical and societal implications of humans cognitively extended by AI. *AIES’19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 507–513.
- Howell, R. J. (2016). Extended Virtues and the Boundaries of Persons. *Journal of the American Philosophical Association*, 2(1), 146–163.
- Kirchhoff, M. D., & Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: A Third Wave View*. Routledge.
- Milgram, S. (1974). *Obedience to authority: An Experimental View*. Harpercollins.
- Olson, E. T. (2011). The Extended Self. *Minds and Machines*, 21(4), 481–495.
- Pugh, J., & Pugh, C. (2021). Neurostimulation, doping, and the spirit of sport. *Neuroethics*, 14, 141–158.
- Skorburg, J. A. (2019). Where are virtues? *Philosophical Studies*, 176(9), 2331–2349.
- Slaby, J., & Gallagher, S. (2015). Critical Neuroscience and Socially Extended Minds. *Theory, Culture and Society*, 32(1), 33–59.
- Sneddon, A. (2011). *Like-Minded: Externalism and Moral Psychology*. MIT Press.
- Snow, N. E. (2019). Proliferating Virtues: A Clear and Present Danger? In E. Grimi, J. Haldane, M. M. Mauri Alvarez, M. Wladika, M. Damonte, M. Slote, R. Curren, C. B. Miller, L. Zyl, C. D. Owens, S. J. Roniger, M. Mangini, N. Snow, & C. Toner (Eds.), *Virtue Ethics: Retrospect and Prospect* (pp. 177–196). Springer.
- Wedgwood, R. (2006). The normative force of reasoning. *Noûs*, 40(4), 660–686.