Confucian Philosophy for the Digital Good Life

Sta. Maria, J.E.

DOI
10.4233/uuid:84f589b0-87f4-4bce-8446-ab80e40e1be8

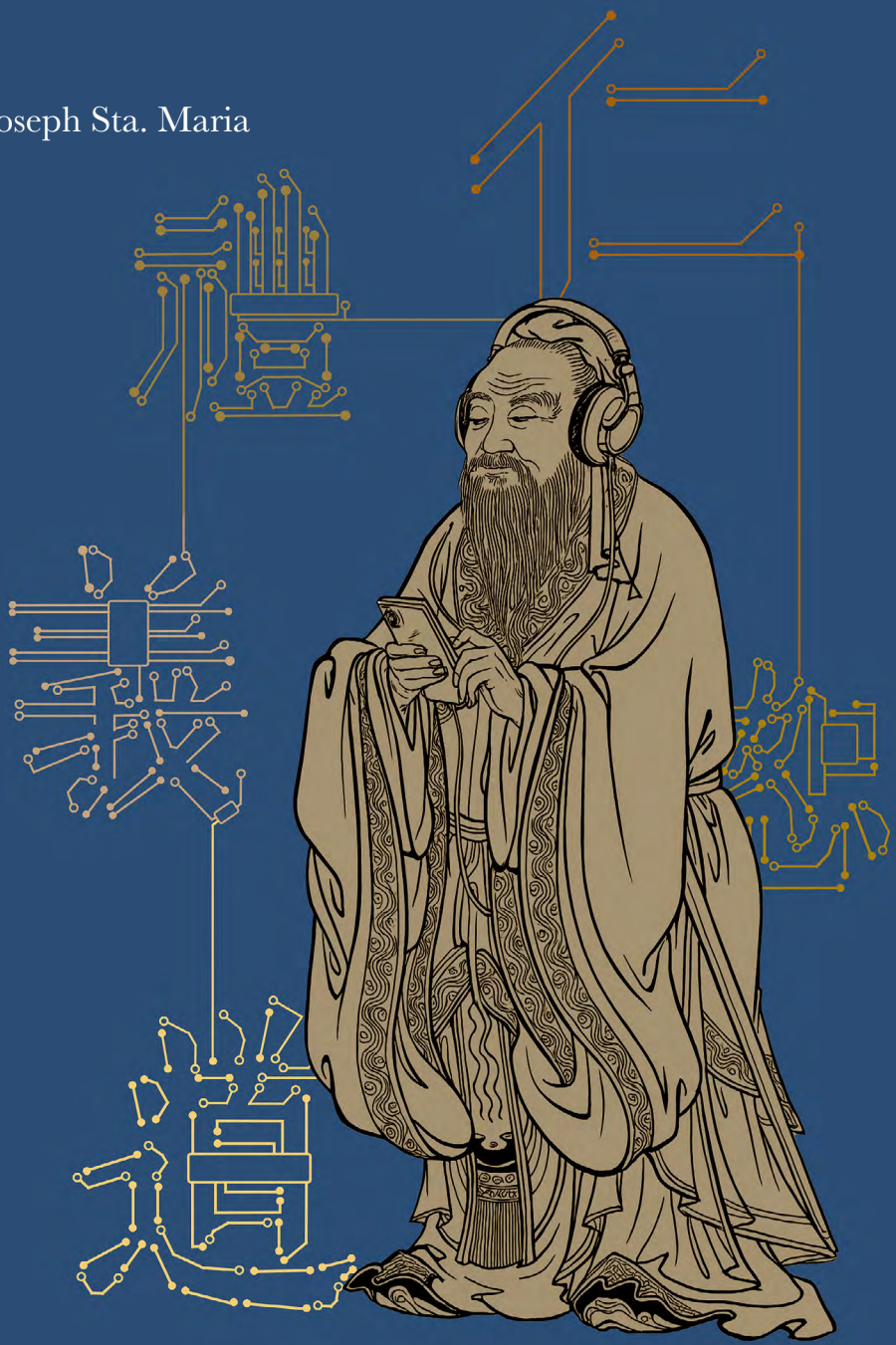
Publication date
2025

Document Version
Final published version

Citation (APA)
Sta. Maria, J. E. (2025). *Confucian Philosophy for the Digital Good Life*. [Dissertation (TU Delft), Delft University of Technology]. 4TU.Centre for Ethics and Technology. https://doi.org/10.4233/uuid:84f589b0-87f4-4bce-8446-ab80e40e1be8

Confucian Philosophy for the Digital Good Life

Joseph Sta. Maria



Simon Stevin Series in the Ethics of Technology

Confucian Philosophy for the Digital Good Life

Confucian Philosophy for the Digital Good Life

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of Rector Magnificus Prof. dr. ir. T. H. J. J. van der Hagen,
chair of the Board of Doctorates
to be defended publicly on Monday 1, December 2025, at 12:30 o'clock

by

Joseph Emmanuel De Los Santos STA MARIA
Master of Arts Philosophy, Ateneo De Manila University, Philippines
Born in Manila, Philippines

This dissertation has been approved by the promotors.

Composition of doctoral committee:

Rector Magnificus	chairperson
Prof. dr. S. Roeser	Delft University of Technology, promotor
Dr. E. Ziliotti	Delft University of Technology, copromotor

Independent members:

Prof. dr. ir. I.R. van de Poel	Delft University of Technology
Prof. dr. P. A. E. Brey	University of Twente
Prof. dr. D.L. Berger	Leiden University
Dr. P.-H. Wong	Hong Kong Baptist University
Prof. dr. ir. B. Taebi	Delft University of Technology, reserve member

Research for this thesis was made possible by the Netherlands Organization for Scientific Research (NWO).

esdit This work is part of the research programme Ethics of Socially Disruptive Technologies, which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

The Simon Stevin Series in the Ethics of Technology is an initiative of the 4TU Centre for Ethics and Technology. Contact: info@ethicsandtechnology.eu

Designer of Cover Image: Yunxuan Miao
ISBN: 978-94-6384-837-4
ISSN: 1574-941X

Copies of this publication may be ordered from the 4TU.Centre for Ethics and Technology, info@ethicsandtechnology.eu
For more information, see http://www.ethicsandtechnology.eu

Table of Contents

List of Figures

List of Tables

List and Status of Papers Used in the Dissertation

Chapters 2 to 5 of this dissertation consist of four papers that the author has written. All these four papers have been co-authored, with the author of this dissertation being the first author of each. Provided below is information about these four papers, including their publication status and the relative contribution of each author. The contributions are stated in terms of the sections (or subsections) of the chapters in this dissertation.

Chapter 2:
Title of Chapter/Paper:
Counteracting Digital McMindfulness from a Neo-Confucian Perspective

Authors:
Joseph Sta. Maria, Matthew Dennis

Contributions:
Sta. Maria wrote parts of Section 2.1 and wrote Sections 2.2 to 2.7. Dennis wrote parts of Section 2.1, contributed to the paper's structure, and made edits throughout the paper.

Status:
The paper is undergoing external review for inclusion in the anthology *The Future of Digital Wellbeing,* to be published by Amsterdam University Press.

Chapter 3:
Title of Chapter/Paper:
Addressing Online Gaming Toxicity from a Confucian Perspective

Authors:
Joseph Sta. Maria, Elena Ziliotti

Contributions:
Sta. Maria wrote parts of Sections 3.1 and 3.5, and wrote Sections 3.2 through 3.4. Ziliotti wrote parts of Sections 3.1 and 3.5, contributed to the paper's structure, and made edits throughout the paper.

Status:
The paper is published in the *Journal of Confucian Philosophy and Culture.*

Sta. Maria, J., & Ziliotti, E. (2022). Addressing online gaming toxicity from a Confucian perspective. *Journal of Confucian Philosophy and Culture, 38*, 131–152. https://doi.org/10.22916/jcpc.2022..38.131

Chapter 4:
Title of Chapter/Paper:
Addressing Online Gaming Toxicity from a Confucian Perspecive

Authors:
Joseph Sta. Maria, Sabine Roeser

Contributions:
Sta. Maria wrote Sections 4.1 through 4.3 and Sections 4.5 through 4.7. Roeser wrote Section 4.4, contributed to the paper's structure, and made edits throughout the paper.

Status:
The paper has not yet been submitted to a journal.

Chapter 5:
Title of Chapter/Paper:
Conceptualizing Conceptual Resilience for Living Well with AI: Lessons from Confucian Philosophy

Authors:
Joseph Sta. Maria, Samuela Marchiori

Contributions:

This paper is a modified and expanded version of a previous paper titled "Conceptualizing Conceptual Resilience: A Comparative Approach," written by the same authors and in which Marchiori is the first author. For the paper in this dissertation, however, Sta. Maria is the first author. He has written Sections 5.1, 5.4, and 5.6. He has partially written Sections 5.2 and mostly written Sections 5.3 and 5.5. Marchiori has written most of 5.2 and a few parts of 5.3 and 5.5.

Status:

This paper has not yet been submitted to a journal.

Summary

This dissertation aims to explore the normative insights that a Confucian-inspired philosophical approach can offer regarding the pursuit of the good life in relation to digital technologies.

The continuing technological advancements of the past decades have led to the rise of socially disruptive technologies (SDTs), that is, technologies with the potential to alter or challenge the social arrangements and practices that constitute human life. Digital technologies are a particularly salient kind of SDTs due to their prevalence. An example of this is social media. As of 2024, more than 60% of the global population uses social media (DataReportal, 2024). Social media enables instant and continuous interaction between people from across the world. Nevertheless, these interactions can be disembodied – abstracted from physical and social cues and arrangements that conventionally serve as determinants for what a relationship should be like between people. Thus, social media has provided a mode of establishing human relations that arguably disrupts the more physically circumscribed way of human relating, which was the previous norm. Social media can afford other forms of social disruption, such as the disruption of the traditional ways democracy has been practiced (Ziliotti et al., 2023). For instance, the connecting power of social media has enabled democratic conversation about local or national issues to expand to a transnational level. As a result, the traditional "demos," once confined to a specific political community, now includes voices from around the world.

Digital technology's capacity for social disruption can affect people's prospects of living well or achieving the "good life." For example, the immediate, continuous, yet disembodied way social media connects people can lead to unhealthy online relationships, as users can easily establish such relationships without having ample familiarity with the people they are interacting with (Dennis & Ziliotti, 2022). Thus, the question arises as to how human beings can live well or achieve the good life in connection with digital technologies and their ever-increasing influence on human existence. I call this the question of the "digital good life," with the term "digital good life" referring to the idea of attaining or maintaining the good life in relation to digital technologies. The question of how to attain the good life is not new. It can be considered one of the perennial questions in philosophy. The question of how to live a good life in relation to technology, including certain digital technologies, is not

wholly unprecedented either (e.g., Brey et al., 2012). However, what makes this research novel is that it draws on a non-Western philosophy, namely Confucian philosophy, for insights on how to answer the question of the digital good life.

There are two reasons why this dissertation chooses to use Confucian philosophy for the above purpose. The first is that it can address an individualistic bias that dominant theories of (digital) well-being, or the (digital) good life, are in danger of falling into. These theories of well-being, such as hedonistic, desire-based, and objective-list theories, all focus on the individual as the subject of well-being. This approach is in danger of conceptually isolating the individual human being, thereby ignoring the relational aspects of human existence and their role in flourishing. This is especially problematic, considering that digital technologies have dramatically increased the interconnections between human beings and thus the influence that relationships have on the chances of people living well or not. Confucian philosophy has the resources to address this because it portrays human existence and the good life as inherently relational. For instance, in the Confucian classic, the *Great Learning* (*Da Xue* 大學), the good life requires spreading one's virtue to benefit the spheres of relations one is situated in, such as the family, the state, and the world. Conversely, one's possibility of achieving the good life is to an important extent determined by one's environment, such as one's socio-political relations. Thus, a Confucian-inspired approach to the good life might prove adequate in capturing how human relations influence the prospect of living well in this digital world. [1]

The second reason this dissertation chooses to use Confucian philosophy is that non-Western philosophy is unrepresented in the philosophical discussion of the digital good life. The lack of representation echoes the general marginalization of non-Western philosophy in contemporary (academic) philosophy. This is an inequitable state of affairs; therefore, the inclusion of non-Western philosophy is a matter of justice. Moreover, due to the global reach of digital technology, the question of the digital good life also concerns people who are more influenced by non-Western philosophies than their Western counterparts. Finally, including non-Western

---

[1] I am aware that other non-Western philosophical traditions (e.g., Ubuntu, Buddhist philosophy) might also have relational conceptions of human life and flourishing. This does not undermine this dissertation, however, since my claim is merely that Confucian philosophy, with its relational understanding of human life, can be a fitting conceptual resource for approaching the question of the digital good life. It is by no means the only one. Indeed, in the conclusion of this dissertation, I emphasize the need to explore other non-Western philosophical traditions to further enrich the philosophical discourse about the digital good life.

philosophy in the aforementioned discussion may offer insights into answering the question of the digital good life that might be overlooked if it were excluded. An example of this is the aforementioned idea that Confucian philosophy might offer a more relational notion of flourishing that may better align with the digitally interconnected world we inhabit today. Thus, using non-Western philosophies as an approach for pursuing the digital good life might be more amenable to them.

Considering the above, the main research question this dissertation attempts to answer is: "**What normative insights can a Confucian-inspired philosophical approach contribute to the pursuit of the digital good life?**"

To answer the question, the dissertation or research draws inspiration from the *Great Learning* in terms of conceiving the scope of the digital good life. That is, the dissertation considers the subject of the digital good life as encompassing not only the individual human but also the spheres of relations in which they are situated. These spheres of relations – modified from the *Great Learning* to better fit the contemporary context – are the self, the interpersonal, the political, and the world. Using this framework, the research explores four ethical issues arising from distinct digital technologies, each impacting a different sphere. These issues affect users' ability to live well in relation to the digital technologies involved. These issues, the digital technology they are connected to, and the sphere of relations they belong in are as follows:

1. The issue of "**Mcmindfulness**" as affecting **mindfulness apps** in the sphere of the **self.**
2. The issue of **toxic behavior** in **online multiplayer games** in the sphere of the **interpersonal.**
3. The issue of **political polarization** in **social media** in the sphere of the **political.**
4. The issue of **conceptual disruption** afforded by **AI systems** in the sphere of the **world**.

The research draws on the conceptual resources of Confucian philosophy to address these issues, with the aim of helping users achieve the digital good life. The Confucian-inspired answers to the issues will provide *specific insights* into the ways Confucian philosophy can contribute to the pursuit of a digital good life. These insights are called "specific" since they correspond only to a particular issue/technology/sphere. The research will also identify the common ideas among these *specific insights*, thereby

articulating *general insights* that Confucian philosophy can provide regarding the advancement of the digital good life. These two kinds of insights will then serve as the answer to the main research question.

The structure of the dissertation is as follows: The Introduction chapter provides further details on the points mentioned above. Chapters 2-5 then each address one of the aforementioned issues through a Confucian-inspired approach.

Chapter 2 explores the issue known as "McMindfulness," which is associated with mindfulness apps. Mindfulness apps fall within the sphere of the self, as they are used individually and influence an individual's self-perception, as the phenomenon of McMindfulness demonstrates. McMindfulness, in turn, refers to the phenomenon where commodified mindfulness products, such as mindfulness apps, reinforce inequitable and onerous (neoliberal) social structures by implying that the root of people's stress and anxiety is not these aforementioned structures but rather their inability to cope with them through mindfulness practices. By doing this, mindfulness apps reinforce a neoliberal conception of the self, in which the individual is solely responsible for their own well-being. Moreover, this self must continually seek self-improvement, in terms of productivity and ability to cope with burdensome neoliberal social arrangements, because it is in constant competition with others. Chapter 2 explores how Confucian philosophy can serve to counter McMindfulness and its effects. In particular, the chapter suggests Confucian-inspired design features in mindfulness apps that can help users adopt and enact a more relational and ethical conception of the self and human flourishing, one where well-being or the good life can be achieved by modifying one's social environment and collaborating with others to do so.

Chapter 3 delves into the issue of toxic behavior as prevalent in online multiplayer games. These games operate at the interpersonal level, as they mediate the relationships between players. Toxic behavior in these games refers to abusive ways of interacting with other players, ranging from verbal insults to making one's team lose intentionally. The chapter examines a specific game, *League of Legends*, which is particularly notorious for its toxic behavior. The chapter assesses that, among other things, one of the main contributors to toxic behavior in *League* is its individual ranking system coupled with its team-based gameplay. That is, players are rewarded by individually gaining higher ranks as they win more matches. However, these matches are played with teams. Moreover, these teams change from one match to another. The effect of this is that players are more likely to view their teammates as temporary means to achieve their goal of attaining a high rank. This then disposes players to treat

their teammates in a toxic way. The chapter suggests Confucian-inspired game features that can minimize, if not eliminate, this toxic behavior. Among these features are those that help inculcate the Confucian values of sympathetic understanding, respect, and interpersonal harmony in these players.

Chapter 4 explores the issue of online political polarization as it exists in social media. By being conducive to political polarization, social media affects the sphere of the political by affecting the social cooperation and cohesion required by a political entity to ensure the well-being of its members (cf. Ziliotti 2022). The chapter assesses that online polarization can be considered largely as "affective polarization" where people are polarized largely due to hostile emotions between partisan groups rather than ideological differences. The chapter, however, argues that, although emotions might be a major driver of polarization, the solution is not to dismiss or eliminate emotions. Building on theories of cognitive emotions, the chapter argues that emotions can be a means of perceiving and appreciating values. This means that emotions can be tapped as a way for people to appreciate values that are inimical to polarization. The chapter then draws inspiration from Confucian philosophy to provide concrete recommendations on how to engage cognitive emotions, thereby helping online users adopt values and attitudes that can reduce polarization. Confucian philosophy is a promising resource for this endeavor because it also assumes the integrated nature of cognition and emotions and considers virtuous emotions as an essential component of flourishing.

Chapter 5 examines the issue of conceptual disruption as afforded by AI systems. Conceptual disruption occurs when concepts cease to fulfill their intended function, such as indicating, classifying, or defining their referents adequately. Conceptual disruption can be brought about by novel states of affairs that obtain due to technologies such as AI systems. For instance, AI "friendships" have the potential to disrupt the concept of friend since they bring into question or challenge what the latter refers to. AI systems belong to the sphere of the world since technology powered by AI systems is present globally and has permeated almost every aspect of human life. Moreover, conceptual disruption affects concepts that, in turn, are the means by which people fundamentally understand and navigate their world. Thus, not only by giving rise to social disruption, but also by affording conceptual disruption, AI systems can influence people's world. Although conceptual disruption is not necessarily bad, it may impede the discourse on AI ethics and governance, as the very concepts used in that discourse can be disrupted. This, in turn, may delay the articulation of ideas or policies intended to safeguard human rights and well-being from the potential

harms of AI. The chapter proposes that a way to handle conceptual disruption is by engineering resilient concepts, that is, concepts that are immune to such disruption. The chapter also proposes two strategies for engineering resilient concepts, namely, conceptual resilience as immutability (CRI) and conceptual resilience as adaptation (CRA). The chapter then draws on Confucian philosophy to provide a concrete example of what engineering for CRA might entail. Specifically, the chapter draws inspiration from Confucius' use of ethical concepts in the *Analects* to sketch the idea of an "indicative concept," the function of which is not to define its referent but only to describe it in a non-universal and non-exhaustive manner, and in a way that can help people to flourish ethically in their specific contexts. Such a concept arguably exhibits CRA because, by not purporting to exhaustively define its referent, it can avoid disruption that might come about due to changes in the definition or meaning of the aforesaid referent. Instead, the indicative concept can flexibly adapt to such changes in meaning.

The concluding chapter of the thesis summarizes the findings of chapters 2-5, thereby identifying the *specific insights* that Confucian philosophy can contribute to the pursuit of the digital good life. The conclusion also identifies five overarching normative insights that are common to the *specific insights*. These are the *general insights*. For lack of space, only the general insights will be enumerated. The first four are (1) the crucial role of pause and reflection, (2) the significance of engaging the cognitive, affective, and sensory faculties of users, as well as (3) "meeting users where they are" when it comes to strategies for advancing the digital good life, and (4) the importance of making users aware of the shared nature of the digital technology and establishing cooperative and ethical relations with other users.

These four insights can be understood as indicating the different parts of Confucian moral self-cultivation, which, from the Confucian perspective, is the process by which people can attain the good life. Coupling this with the idea that these insights involve the use of digital technologies yields a final insight, which is that (5) digital technologies have the potential to support the entire process of moral self-cultivation, through which people can attain the good life. The implication of this answer is optimistic: despite the potential harms of digital technologies, they can nevertheless be a comprehensive means by which we can attain the good life, understood as flourishing together. This answer also invites a positive way of understanding digital technology's potential for social disruption. Pursuing the digital good life might require a disruption of familiar ways of thinking and acting that hinder living well. When designed for self-cultivation, digital technology can facilitate this

disruption and promote flourishing. Thus, the capacity of digital technology for social disruption can be harnessed to help us live well with one another. This, and the rest of the abovementioned insights, both specific and general, thus serve as the answer to the main research question.

The concluding chapter also presents other implications derived from the dissertation, including those highlighted by its limitations. These are, namely, the importance of empirically assessing the Confucian-inspired recommendations given in chapters 2-5, and the equally significant task of exploring other non-Western philosophies and their potential contribution to the philosophical discourse of the digital good life. Finally, the conclusion also states the academic implications and relevance of the research for society by showing how it fulfils the goals of the Ethics of Socially Disruptive Technologies (ESDiT) programme, of which the aforesaid research is a part.

Samenvatting

Deze dissertatie heeft tot doel de normatieve inzichten te verkennen die een door het confucianisme geïnspireerde filosofische benadering kan bieden met betrekking tot het nastreven van een goed leven in relatie tot digitale technologieën.

De voortdurende technologische vooruitgang van de afgelopen decennia heeft geleid tot de opkomst van sociaal ontwrichtende technologieën ('socially disruptive technologies' of SDT's), dat wil zeggen technologieën met het potentieel om de sociale arrangementen en praktijken die het menselijk leven vormen te veranderen of uit te dagen. Digitale technologieën zijn een bijzonder opvallende soort SDT's vanwege hun prevalentie. Een voorbeeld hiervan is sociale media. Vanaf 2024 gebruikt meer dan 60% van de wereldbevolking sociale media (DataReportal, 2024). Sociale media maken directe en continue interactie mogelijk tussen mensen van over de hele wereld. Niettemin kunnen deze interacties 'ontlichaamd' zijn - geabstraheerd van fysieke en sociale signalen en arrangementen die conventioneel dienen als bepalende factoren voor hoe een relatie tussen mensen zou moeten zijn. Zo maken sociale media vormen van menselijke relaties mogelijk die de eerdere norm van menselijke relaties verstoort, waarbij fysiek contact cruciaal was. Sociale media kunnen ook tot andere vormen van sociale ontwrichting leiden, zoals de verstoring van de traditionele manieren waarop democratie wordt beoefend (Ziliotti et al., 2023). Zo heeft de verbindende kracht van sociale media ervoor gezorgd dat democratische discussies over lokale of nationale kwesties zich tot een transnationaal niveau hebben kunnen uitbreiden. Hierdoor omvat de traditionele 'demos', ooit beperkt tot een specifieke politieke gemeenschap, nu stemmen van over de hele wereld.

Het vermogen van digitale technologie om maatschappelijke ontwrichting teweeg te brengen, kan de vooruitzichten van mensen om een 'goed leven' te leiden of te bereiken, beïnvloeden. De directe, continue, maar toch onstoffelijke manier waarop sociale media mensen met elkaar verbinden, kan bijvoorbeeld leiden tot ongezonde online relaties, omdat gebruikers dergelijke relaties gemakkelijk kunnen opbouwen zonder al te veel vertrouwdheid met de mensen met wie ze omgaan (Dennis & Ziliotti, 2022). De vraag rijst dan ook hoe mensen een goed leven kunnen leiden of bereiken in relatie tot digitale technologieën en hun steeds toenemende invloed op het menselijk bestaan. Ik noem dit de vraag van het 'digitale goede leven', waarbij de term 'digitaal

goed leven' verwijst naar het idee om een goed leven te bereiken of te behouden in relatie tot digitale technologieën. De vraag hoe een goed leven te bereiken is niet nieuw. Ze kan worden beschouwd als een van de eeuwige vragen in de filosofie. De vraag hoe een goed leven te leiden in relatie tot technologie, inclusief bepaalde digitale technologieën, is ook niet geheel ongekend (bijv. Brey et al., 2012). Wat dit promotieonderzoek bijzonder maakt, is dat het een niet-westers filosofisch kader gebruikt, namelijk de confucianistische filosofie, om inzicht te verschaffen in de vraag naar het digitale goede leven.

Er zijn twee redenen waarom dit proefschrift ervoor kiest om de confucianistische filosofie voor bovenstaand doel te gebruiken. Ten eerste kan het een alternatief bieden voor de individualistische 'bias' waar dominante theorieën over (digitaal) welzijn, of het (digitale) goede leven, in dreigen te vervallen. Deze theorieën over welzijn, zoals de hedonistische, op verlangens gebaseerde en objectieve-lijsttheorieën richten zich allemaal op het individu als het onderwerp van welzijn. Deze benaderingen dreigen de individuele mens conceptueel te isoleren en daarmee de relationele aspecten van het menselijk bestaan en het goede leven te negeren. Dit is met name problematisch, aangezien digitale technologieën de onderlinge verbindingen tussen mensen en daarmee de invloed van relaties op de kansen van mensen om wel of niet goed te leven drastisch hebben vergroot. De confucianistische filosofie heeft de middelen om dit aan te pakken, omdat zij het menselijk bestaan en het goede leven als inherent relationeel afschildert. In de confucianistische klassieker de *'Grote Leer'* (*Da Xue* 大學) vereist het goede leven bijvoorbeeld het verspreiden van iemands deugdzaamheid ten gunste van de relaties waarin men zich bevindt, zoals het gezin, de staat en de wereld. Omgekeerd wordt iemands mogelijkheid om een goed leven te bereiken in belangrijke mate bepaald door iemands omgeving, zoals iemands sociaal-politieke verhoudingen. Dus een door het confucianisme geïnspireerde benadering van het goede leven zou toereikend kunnen blijken in het vastleggen van hoe menselijke relaties het vooruitzicht van goed leven in deze digitale wereld beïnvloeden.[2]

---

[2] Ik ben me ervan bewust dat andere niet-westerse filosofische tradities (bijvoorbeeld Ubuntu, boeddhistische filosofie) mogelijk ook relationele opvattingen over het menselijk leven en de bloei ervan hebben. Dit ondermijnt dit proefschrift echter niet, aangezien mijn bewering slechts is dat de confucianistische filosofie, met haar relationele begrip van het menselijk leven, een passende conceptuele bron kan zijn voor de benadering van de vraag naar het digitale goede leven. Het is zeker niet de enige. Sterker nog, in de conclusie van dit proefschrift benadruk ik de noodzaak om andere niet-westerse filosofische tradities te verkennen om het filosofische discours over het digitale goede leven verder te verrijken.

De tweede reden waarom dit proefschrift kiest voor de confucianistische filosofie, is dat niet-westerse filosofie vaak niet vertegenwoordigd is in de filosofische discussie over het digitale goede leven. Dit gebrek aan representatie weerspiegelt de algemene marginalisering van niet-westerse filosofie in de hedendaagse (academische) filosofie. Dit is een onrechtvaardige situatie; daarom is de opname van niet-westerse filosofie een kwestie van rechtvaardigheid. Bovendien, vanwege het wereldwijde bereik van digitale technologie, gaat de vraag naar het digitale goede leven ook over mensen die meer beïnvloed worden door niet-westerse filosofieën dan hun westerse tegenhangers. Ten slotte kan het opnemen van niet-westerse filosofie in de bovengenoemde discussie inzichten bieden in het beantwoorden van de vraag naar het digitale goede leven, die mogelijk over het hoofd zouden worden gezien als deze zou worden uitgesloten. Een voorbeeld hiervan is het eerdergenoemde idee dat de confucianistische filosofie een meer relationeel begrip van bloei zou kunnen bieden dat beter aansluit bij de digitaal verbonden wereld waarin we vandaag de dag leven. Een niet-westerse filosofie zoals het confucianisme kan dan ook een belangrijke bijdrage leveren aan inzichten over hoe het digitale goede leven te bereiken.

Gezien het bovenstaande is de belangrijkste onderzoeksvraag die dit proefschrift probeert te beantwoorden: "**Welke normatieve inzichten kan een door het confucianisme geïnspireerde filosofische benadering bijdragen aan het nastreven van het digitale goede leven?**"

Om deze vraag te beantwoorden, put het proefschrift of onderzoek inspiratie uit de *Grote Leer* voor het begrip van de reikwijdte van het digitale goede leven. Dat wil zeggen, het proefschrift beschouwt als het onderwerp van het digitale goede leven niet alleen individuen, maar ook de relationele sferen waarin zij zich bevinden. Deze relationele sferen –geïnspireerd op de *Grote Leer* maar in aangepaste vorm om beter aan te sluiten bij de hedendaagse context – zijn het zelf, het interpersoonlijke, het politieke en de wereld. Met behulp van dit kader onderzoekt het onderzoek vier ethische kwesties die voortvloeien uit verschillende digitale technologieën, die elk een andere sfeer beïnvloeden. Deze kwesties beïnvloeden het vermogen van gebruikers om goed te leven in relatie tot de betrokken digitale technologieën. Deze kwesties, de digitale technologie waarmee ze verbonden zijn en de relationele sferen waartoe ze behoren, zijn als volgt:

1. De kwestie van "**Mcmindfulness**" en de invloed ervan op **mindfulness-apps** op het vlak van het **zelf.**

2. Het probleem van **toxisch gedrag** in **online multiplayer games** op **interpersoonlijke vlak.**
3. Het probleem van **politieke polarisatie** in **sociale media** in de sfeer van het **politieke.**
4. Het probleem van **conceptuele verstoringen die AI-systemen** in de **wereld** teweegbrengen.

Het onderzoek maakt gebruik van de conceptuele bronnen van de confucianistische filosofie om deze kwesties aan te pakken, met als doel om inzichten te verwerven die gebruikers kunnen helpen een digitaal goed leven te bereiken. De door het confucianisme geïnspireerde antwoorden op de kwesties zullen *specifieke inzichten* verschaffen in de manieren waarop de confuciaanse filosofie kan bijdragen aan het nastreven van een digitaal goed leven. Deze inzichten worden 'specifiek' genoemd omdat ze alleen betrekking hebben op een specifiek probleem/technologie/domein. Op basis van deze *specifieke inzichten* leidt het onderzoek ook tot *algemene inzichten* met betrekking tot hoe de confucianistische filosofie kan bijdragen aan de bevordering van een digitaal goed leven. Deze twee soorten inzichten dienen vervolgens als antwoord op de hoofdvraag van het onderzoek.

De structuur van het proefschrift is als volgt: het inleidende hoofdstuk geeft verdere details over de bovengenoemde punten. Hoofdstukken 2–5 behandelen vervolgens elk een van de eerder genoemde kwesties via een door het confucianisme geïnspireerde benadering.

Hoofdstuk 2 onderzoekt het onderwerp 'McMindfulness', dat verband houdt met mindfulness-apps. Mindfulness-apps vallen binnen de sfeer van het zelf, omdat ze individueel worden gebruikt en de zelfperceptie van een individu beïnvloeden, zoals het fenomeen McMindfulness laat zien. McMindfulness verwijst op zijn beurt naar het fenomeen waarbij gecommercialiseerde mindfulnessproducten, zoals mindfulness-apps, onrechtvaardige en belastende (neoliberale) sociale structuren versterken door te impliceren dat de oorzaak van de stress en angst van mensen niet deze bovengenoemde structuren zijn, maar eerder hun onvermogen om ermee om te gaan, waarvoor mindfulness-oefeningen als remedie worden gepresenteerd. Door dit te doen, versterken mindfulness-apps een neoliberaal zelfbeeld, waarin het individu als enige verantwoordelijk is voor zijn eigen welzijn. Bovendien moet dit zelf voortdurend streven naar zelfverbetering, in termen van productiviteit en het vermogen om met belastende neoliberale sociale arrangementen om te gaan, omdat het in constante concurrentie met anderen verkeert. Hoofdstuk 2 onderzoekt hoe de

confucianistische filosofie kan dienen als tegenwicht tegen McMindfulness en de effecten ervan. In het hoofdstuk worden met name ontwerpkenmerken van het confucianisme in mindfulness-apps voorgesteld die gebruikers kunnen helpen een meer relationeel en ethisch concept van het zelf en menselijke bloei te omarmen en uit te voeren. Een concept waarin welzijn of een goed leven kan worden bereikt door de sociale omgeving aan te passen en door samen te werken met anderen om dit te bereiken.

Hoofdstuk 3 gaat dieper in op de kwestie van toxisch gedrag zoals dat veel voorkomt in online multiplayergames. Deze games opereren op interpersoonlijk niveau, omdat ze de relaties tussen spelers bemiddelen. Toxisch gedrag in deze games verwijst naar beledigende manieren van omgaan met andere spelers, variërend van verbale beledigingen tot het opzettelijk laten verliezen van je team. Het hoofdstuk onderzoekt een specifieke game, *League of Legends*, die bijzonder berucht is om zijn toxisch gedrag. Het hoofdstuk beargumenteert dat een van de belangrijkste factoren voor toxisch gedrag in *League* het individuele rangsysteem in combinatie met de teamgebaseerde gameplay is. Dat wil zeggen dat spelers worden beloond door individueel hogere rangen te behalen naarmate ze meer wedstrijden winnen. Deze wedstrijden worden echter met teams gespeeld. Bovendien wisselen deze teams van wedstrijd tot wedstrijd. Het effect hiervan is dat spelers hun teamgenoten eerder zien als tijdelijke middelen om hun doel van een hoge rang te bereiken. Dit leidt er vervolgens toe dat spelers hun teamgenoten op een toxisch manier behandelen. In dit hoofdstuk worden door het confucianisme geïnspireerde spelkenmerken voorgesteld die dit toxische gedrag kunnen minimaliseren, zo niet elimineren. Deze kenmerken dragen bij aan het bijbrengen van de Confucianistische waarden van begrip, respect en interpersoonlijke harmonie bij deze spelers.

Hoofdstuk 4 onderzoekt de kwestie van online politieke polarisatie zoals die bestaat in sociale media. Door politieke polarisatie te bevorderen, beïnvloeden sociale media de politieke sfeer door hun impact op de sociale samenwerking en cohesie die een politieke entiteit nodig heeft om het welzijn van haar leden te waarborgen (vgl. Ziliotti 2022). Het hoofdstuk beargumenteert dat online polarisatie grotendeels kan worden beschouwd als "affectieve polarisatie", waarbij mensen grotendeels gepolariseerd raken door vijandige emoties tussen groepen waardoor ideologische verschillen worden uitvergroot. Het hoofdstuk betoogt echter dat, hoewel emoties een belangrijke drijfveer voor polarisatie kunnen zijn, de oplossing niet ligt in het negeren of elimineren van emoties. Voortbouwend op theorieën over cognitieve emoties, stelt het hoofdstuk dat emoties een middel kunnen zijn om aandacht voor waarden waar

te hebben. Dit betekent dat emoties een rol kunnen spelen voor aandacht voor waarden die polarisatie kunnen tegengaan. Het hoofdstuk put vervolgens inspiratie uit de confucianistische filosofie om concrete aanbevelingen te doen over hoe cognitieve emoties kunnen worden ingezet, waardoor online gebruikers waarden en attitudes kunnen aannemen die polarisatie kunnen verminderen. De Confucianistische filosofie is een veelbelovende bron voor dit streven, omdat deze filosofie ook uitgaat van de geïntegreerde aard van cognitie en emoties, en deugdzame emoties beschouwt als een essentieel onderdeel van menselijke morele ontwikkeling.

Hoofdstuk 5 onderzoekt de kwestie van conceptuele verstoring zoals die door AI-systemen wordt veroorzaakt. Conceptuele verstoring treedt op wanneer concepten niet langer hun beoogde functie vervullen, zoals het adequaat aangeven, classificeren of definiëren van datgene waaraan ze refereren (de 'refererent'). Conceptuele verstoring kan worden veroorzaakt door nieuwe ontwikkelingen die ontstaan dankzij technologieën zoals AI-systemen. AI-"vriendschappen" hebben bijvoorbeeld de potentie om het concept van een vriend te verstoren, omdat ze de betekenis ervan in twijfel trekken of uitdagen. AI-systemen behoren tot de wereld, aangezien technologie die door AI-systemen wordt aangestuurd wereldwijd aanwezig is en vrijwel elk aspect van het menselijk leven heeft doordrongen. Bovendien beïnvloedt conceptuele verstoring concepten, die op hun beurt de middelen zijn waarmee mensen hun wereld fundamenteel begrijpen en navigeren. Dus, niet alleen door maatschappelijke verstoring te veroorzaken, maar ook door conceptuele verstoring te veroorzaken, kunnen AI-systemen de wereld van mensen beïnvloeden. Hoewel conceptuele verstoring niet per se slecht is, kan het het discours over AI-ethiek en -bestuur belemmeren, omdat juist de concepten die in dat discours worden gebruikt, verstoord kunnen raken. Dit kan op zijn beurt de formulering van ideeën of beleidsmaatregelen vertragen die bedoeld zijn om mensenrechten en welzijn te beschermen tegen de potentiële schade van AI. Het hoofdstuk stelt voor om conceptuele verstoring aan te pakken door veerkrachtige concepten te ontwikkelen, dat wil zeggen concepten die immuun zijn voor dergelijke verstoring. Het hoofdstuk stelt ook twee strategieën voor om veerkrachtige concepten te ontwikkelen, namelijk conceptuele veerkracht als onveranderlijkheid (CRI) en conceptuele veerkracht als aanpassing (CRA). Het hoofdstuk baseert zich vervolgens op de confucianistische filosofie om een concreet voorbeeld te geven van wat ontwikkeling voor CRA zou kunnen inhouden. Het hoofdstuk put met name inspiratie uit Confucius' gebruik van ethische concepten in de *Analecten* om het idee van een 'indicatiefconcept' te schetsen, waarvan de functie niet is om de referent te definiëren, maar slechts om het op een niet-universele en niet-

uitputtende manier te beschrijven, en op een manier die mensen kan helpen om ethisch te floreren in hun specifieke contexten. Een dergelijk concept vertoont aantoonbaar CRA omdat het, door niet te pretenderen de referent uitputtend te definiëren, verstoringen kan vermijden die zouden kunnen ontstaan door veranderingen in de definitie of betekenis van de genoemde referent. In plaats daarvan kan het indicatief concept zich flexibel aanpassen aan dergelijke betekenis-veranderingen.

Het afsluitende hoofdstuk van het proefschrift vat de bevindingen van hoofdstukken 2-5 samen en identificeert daarmee de *specifieke inzichten* die de confucianistische filosofie kan bijdragen aan het nastreven van het digitale goede leven. De conclusie identificeert tevens vijf overkoepelende normatieve inzichten op basis van de *specifieke inzichten*. Dit zijn de *algemene inzichten*. Vanwege ruimtegebrek worden alleen de algemene inzichten opgesomd. De eerste vier zijn (1) de cruciale rol van pauze en reflectie, (2) het belang van het betrekken van de cognitieve, affectieve en sensorische vermogens van gebruikers, evenals (3) "gebruikers ontmoeten waar ze zijn" als het gaat om strategieën voor het bevorderen van het digitale goede leven, en (4) het belang van het bewustmaken van gebruikers van de gedeelde aard van de digitale technologie en het aangaan van coöperatieve en ethische relaties met andere gebruikers.

Deze vier inzichten kunnen worden opgevat als een indicatie van de verschillende onderdelen van confucianistische morele zelfcultivatie, wat vanuit confucianistisch perspectief het proces is waarmee mensen het goede leven kunnen bereiken. Door dit te combineren met het idee dat deze inzichten het gebruik van digitale technologieën inhouden, ontstaat een laatste inzicht, namelijk dat (5) digitale technologieën de potentie hebben om het gehele proces van morele zelfcultivatie te ondersteunen, waardoor mensen het goede leven kunnen bereiken. De implicatie van dit antwoord is optimistisch: ondanks de potentiële nadelen van digitale technologieën kunnen ze desalniettemin een allesomvattend middel zijn waarmee mensen het goede leven kunnen bereiken, opgevat als samen 'floreren'. Dit antwoord nodigt ook uit tot een positieve benadering van het potentieel van digitale technologie voor maatschappelijke ontwrichting. Het nastreven van het digitale goede leven vereist mogelijk een ontwrichting van vertrouwde manieren van denken en handelen die een goed leven belemmeren. Wanneer digitale technologie is ontworpen voor zelfcultivatie, kan ze deze ontwrichting faciliteren en floreren bevorderen. Zo kan het vermogen van digitale technologie tot maatschappelijke ontwrichting worden benut om ons te helpen goed met elkaar samen te leven. Dit en de overige hierboven

genoemde inzichten, zowel specifiek als algemeen, vormen het antwoord op de hoofdvraag van het onderzoek.

Het afsluitende hoofdstuk presenteert ook andere inzichten die voortvloeien uit het proefschrift, evenals beperkingen. Deze andere inzichten en beperkingen betreffen het belang van een empirische beoordeling van de door confucianisme geïnspireerde aanbevelingen in hoofdstukken 2-5, en de even belangrijke taak om andere niet-westerse filosofieën en hun potentiële bijdrage aan het filosofische discours over het digitale goede leven te verkennen. Ten slotte bespreekt de conclusie ook de filosofische implicaties en de relevantie van het onderzoek voor de samenleving door te laten zien hoe het voldoet aan de doelstellingen van het programma Ethics of Socially Disruptive Technologies (ESDiT), waarvan het bovengenoemde onderzoek deel uitmaakt.

Chapter 1. Introduction

Digital technologies permeate virtually all facets of human life. This has become particularly pronounced during the coronavirus pandemic, in which digital technologies have largely mediated human interaction and work. For instance, schools had to conduct classes entirely through digital communication platforms such as *Zoom* and *Microsoft Teams* because of the impossibility of physical meetings during lockdowns. To be sure, some of the aspects of these digitally mediated activities have been frowned upon. For example, online classes during the pandemic have been considered less conducive to learning when compared to physical classes (West, 2023). Nevertheless, the advantages of these digitally mediated activities have also been acknowledged. For example, working from home through digital means has increased the subjective well-being of women who have to deal with long commutes when working offline (Kroesen, 2022). Working from home through digital means has also substantially lessened vehicle emissions (Santos & Azhari, 2022). Thus, the pandemic has brought into greater relief the considerable impact, both positive and negative, that digital technologies have on human life. Consequently, the importance of maintaining the positive impacts while avoiding the negative ones should now become clearer. This is especially considering that, despite the cessation of the pandemic, the use of digital technologies shows no sign of disappearing anytime soon.

However, technologies "are not neutral intermediaries but actively co-shape people's being in the world: their perceptions and actions, experience and existence" (Verbeek, 2011, p.8). This means digital technologies are not simply tools that help people achieve their goals. Instead, they contribute to the shaping of people's outlooks, behaviors, and their very lives. Moreover, they can disrupt previously existing practices and ideas by which people understand and navigate the world. For instance, the concept and practice of democracy have arguably been challenged by social media, as it has expanded the notion of the people and the public sphere from a local, nation-state-centered arena to a more global field (Ziliotti et al., 2023, pp. 34–52). This capacity for social and conceptual disruption has also increased digital technology's potential to positively or negatively affect people and their chances of living well. Again, social media can be an example. Social media challenges the conventional way human beings develop relationships, i.e., in a gradual and localized manner, by allowing them to immediately interact with a potentially unlimited number of unknown people. This setup increases the likelihood that social media users

will engage in social relations unreflectively and in ways that may ultimately prove detrimental to their well-being (Dennis & Ziliotti, 2022).

Because of the above considerations, there has been an increased interest among scholars from different fields in understanding how digital technologies precisely influence people's chances of living well. This can be considered the issue or question of the **"digital good life." That is to say, it is the question of how to attain or maintain the good life while using, experiencing, and being influenced by digital technologies.**

From a philosophical perspective, the question of the digital good life can be considered an extension of the perennial philosophical question of how to live well. Since the time of the Ancient Greeks, philosophers have inquired about what constitutes a good (human) life and how this kind of life could be achieved (Mendie & Udofia, 2020). Today's philosophers are also asking how people can attain the good life in the face of technologies, including digital technologies, and their growing impact on human existence (e.g., Higgs et al., 2000; Wong, 2010; Brey et al., 2012; Vallor, 2016). At least in the Western philosophical scene, there are three dominant "theories of well-being" that are likely to be used to answer the question of the digital good life. These are generally referred to as the hedonistic, desire-based, and objective-list theories (Parfit, 1984; Haybron, 2008; Kim, 2020; Brey et al., 2012).[3]

Without denying the value of these theories, there are nevertheless a couple of concerns about them that warrant attention. The first is that these theories, and others like them, might focus too much on the individual's well-being or interest. This, in turn, might make them ill-suited for addressing the question of the digital good life since digital technologies have radically intertwined people and their prospects of flourishing, thereby necessitating a more relational view of living well with these technologies. This will be discussed in section 1.2. The second is that these theories reflect the dominance of Western philosophical traditions when it comes to the resources available for addressing the question of the digital good life. This has both practical and ethical implications, which will be explained in Section 1.3. Although these two concerns are distinct, this dissertation shall attempt to address both by drawing on Confucian philosophy as an inspiration or conceptual resource for

[3] The notion that these three kinds of theories are likely to be used in answering the question of the digital good life is supported by the fact that two influential publications about digital well-being/good life, i.e., *The Good Life in a Technological Age* (2013) and *Ethics of Digital Well being*: A Multidisciplinary Approach (2020), identify precisely only these three theories as the major philosophical theories of well-being.

approaching the question of the digital good life. That is, the main research question of this dissertation is **"What normative insights can a Confucian-inspired philosophical approach contribute to the pursuit of the digital good life?"** Section 1.4 will explain why using Confucian philosophy can address the two concerns or gaps presented in sections 1.2 and 1.3. In short, drawing inspiration from Confucian philosophy can address the gap presented in section 1.2 because it has a relational and other-regarding vision of the good life that can counterbalance the individualistic bias of the aforesaid dominant theories of well-being, and aligns more closely with the interconnected nature of human life and flourishing afforded by digital technologies. On the other hand, utilizing Confucian philosophy can address the gap stated in section 1.3, as it is a non-Western philosophy that remains influential in certain areas of the world. Section 1.5 will explain this dissertation's methodology, or how exactly Confucian philosophy will be used to answer the main research question. Finally, section 1.6 will provide an outline of the dissertation by briefly describing its chapters and stating the research subquestions. In the immediately next section, however, I will first clarify the idea of the "digital good life."

1.1. Clarifying the "Digital Good Life"

In this section, I will provide a few preliminary clarifications regarding the conception of the digital good life used in this dissertation. First, let me define digital technologies. I consider digital technologies as those technologies "associated with computations built on discrete binary operations" (Markham, 2020, p.24). Digital technologies are legion, ranging from personal smartphones to office printers to military drones. It is the considerable impact of these technologies on human life that has led philosophers to ask the question of the digital good life.

As for my notion of the digital good life, I intend it to be a philosophical idea, in line with the nature of this dissertation as a research project in philosophy. Thus, I will not explore the literature on the good life and, more particularly, well-being that belongs to other fields such as psychology and economics. Moreover, I do not intend to provide a substantive and definitive account of the digital good life. The digital good life is essentially a subset of the good life; it is the good life as it manifests in people's use and experience of digital technologies. However, what exactly constitutes the good life, particularly its substantive content, has been a matter of debate since the beginning of philosophy. Moreover, the globalized nature of our contemporary world has made it increasingly evident that people from different parts of the world

have varying conceptions of the good life. It remains to be seen whether this plurality of views can be conclusively unified into a single substantive account. Even if such an account is feasible, it might not be desirable since it can eliminate the distinctiveness of the various conceptions of the good life. Taking these into consideration, it would seem imprudent for me to exhaustively and definitively identify the substantive content of the good life, and by extension, the digital good life.

Nevertheless, my idea of the digital good life is generally grounded in the dominant normative perspective that our current global community adheres to (at least formally) and which is centered on the notion "that all humans are equal in rights and dignity" (Hongladarom, 2021, p. 128).[4] Thus, it assumes respect for such dignity and rights. It

---

[4] This move is inspired by Soraj Hongladarom's interpretation of Charles Ess' Pros Hen ethical pluralism (Hongladarom, 2021). *Pros Hen* ethical pluralism is an approach Ess applies in intercultural information and digital ethics (Ess, 2006, 2021). The basic idea is that ethical differences between cultures can be viewed as irreducibly different interpretations or senses of a common set of ideas. *Pros Hen* pluralism thus attempts to find a middle way between an ethical relativism that denies any common ground between diverse ethical traditions and an ethical monism that collapses ethical differences into an identity. Hongladarom attempts to make Pros Hen pluralism more grounded in current social reality by suggesting that the aforesaid common set of ideas should be the normative notions and values that prevail in the overall socio-cultural context in which the Pros Hen approach is being applied (Hongladarom, 2021). This can make the Pros Hen approach compatible with and applicable to the actual experience of contemporary people and societies. This also has the pragmatic advantage of avoiding abstract and potentially interminable metaethical debates about which set of goods or values is objectively true, and what would be its justification.

Hongladarom claims that the globally prevailing normative framework is the one that assumes the equality of human dignity and human rights. This may sound like a "capitulation" to the hegemony of Western values. However, Hongladarom's point is that looking at Western and non-Western cultures from a wholly theoretical perspective might give the impression that these cultures are more isolated than they really are. It may prevent researchers from realizing that, "on the ground," some "Western" values have already been very much adopted by non-Western peoples. Hongladarom supports this by pointing to numerous international agreements that several nations have subscribed to that assume human rights and dignity. Hongladarom also points to his country of Thailand, whose younger generation has, of its own volition, begun advocating for a more egalitarian society based on the aforementioned equality of human dignity and rights (2021). While I do not wholly agree with Hongladarom's interpretation of *Pros Hen* pluralism as it seems to skirt too near to moral relativism, I agree that the notion of universal human dignity and rights is already accepted globally. It serves as the cornerstone of international legal and moral frameworks. It is also subscribed to by most people worldwide; 71% of the global population average considers human rights as reflecting values they believe in, and 72% believe that human rights are a force for good (Griffiths, 2023, p. 7). To be sure, there have been calls to make human rights more accommodating of cultural differences, as seen in the Bangkok Declaration (United Nations, 1993) and in the work of scholars (e.g., Macioce, 2016; Wong, 2020). However, these are not calls to eliminate human rights but to interpret and complement them in light of culturally specific values.

also assumes pluralism or respect for different ways of life. Finally, it assumes the desirability of people achieving well-being while living in harmony, or at least peace, with one another. There are a couple of further features of the digital good life that I will discuss in the following sections (sections 1.2 and 1.4). One is the distinction between digital good life and digital well-being. However, this is connected to the individualistic bias of the aforesaid dominant theories of well-being. Thus, I will discuss these together in the next section.

1.2. Good Life vs Well-Being, and the Individualistic Bias of the Theories of Well-Being

My idea of the digital good life is distinct from digital well-being. Although "good life" and "well-being" are often used interchangeably in the philosophical literature (Kim, 2020), there are two reasons or motivations why I choose to use the former over the latter (i.e., "digital good life" instead of "digital well-being").[5]

First, the term "well-being," including "digital well-being," is often used in other fields such as psychology and economics (Brey, 2012). On the other hand, the "good life," perhaps due to its more normative connotation, is used particularly in ethics and philosophy. Thus, I choose the term "good life" to clarify that this research is philosophical in nature.

The second reason is based on a distinction between the terms "good life" and "well-being" that philosophers have identified (e.g., Swanton, 2003). This distinction turns on the idea that exercising morality can endanger or worsen a person's (non-moral) welfare. Conversely, people can live well-off lives, sometimes by being immoral. Daniel Haybron's response to this tension is to define "well-being" as constituted by the set of things that (non-morally or prudentially) "benefits a person, is in her interest, is good for her, or makes her life go well for her" (Haybron, 2008, p. 29). On the other hand, Haybron defines the "good life" as including well-being *and* any other element that makes a life admirable or "choiceworthy" (2008, 36). This includes morality or virtue, which, though estimable, might not be easily categorized as part of well-being because of how it might bring (non-moral or prudential) harm to its possessor. In other words, the good life encompasses "all the things that ultimately matter in life, whether they benefit the agent or not" (Haybron, 2008, p. 36).

---

[5] Other terms that are interchangeable with "good life" and "well-being" in the philosophical literature are "happiness" and "flourishing" (Kim, 2020).

I choose the term "good life" as Haybron defines it, as it has the advantage of encompassing well-being while also doing justice to the possibility that living well involves acting in ways that benefit others even when these acts do not necessarily benefit oneself. This avoids reducing the idea of the good life to self-interest and makes it a genuinely moral idea.[6] Moreover, and more significantly for this dissertation, I wish to reject the notion that living well is merely a matter of benefiting oneself because it potentially promotes an individualistic conception of the good life. This "individualistic bias" is not appropriate for a vision of the digital good life, as people – and their prospects of flourishing – have become increasingly interconnected due to these technologies. To put it differently, digital technologies have connected people in ways that personal actions previously thought harmless can now significantly impact others and their prospects of flourishing.

For instance, Pak Hang-Wong has argued that AI systems have brought about an "interconnectedness condition" that has expanded one's scope of moral responsibility beyond what is commonly thought (2021). To illustrate, he cites a hypothetical case where Jack, a member of an online job board, affects another member, Bob, and his chances of applying for jobs because the algorithm has judged that both have similar profiles and therefore gives them similar job vacancy ads. Because of this, Wong argues, Jack should inform himself about the general ways the algorithm can make his search results affect the search results of others like Bob. In other words, regarding his search decisions, Jack should not only think for himself but also on behalf of a group, i.e., the group of people with similar profiles to his. Failure to do so might lead to search decisions that could detrimentally affect Bob's chances of getting a good job vacancy ad (e.g., Jack searches for "easy" jobs below his competencies, and this ends up in Bob getting the same job ads below his competencies, which prove useless for Bob). Jack, therefore, arguably has a responsibility to others that would not exist if the AI system of the online job board had not connected him with other users (i.e., had not brought about the interconnectedness condition).

A theory of the good life or well-being that focuses solely on the individual agent and its benefits may not be able to capture the interconnectedness condition, as

---

[6] Nevertheless, adopting Haybron's understanding does not reduce the digital good life to the ethically lived life, i.e., a conception of life that only focuses on moral elements or goods. Haybron still allows the inclusion of non-moral goods in the good life. I follow him in this since good life theories – including the implicit vision of the good life in Confucian philosophy, as will be shown in section 1.4 – typically include non-moral goods, and I have no reason to diverge from this. Examples of such theories include Aristotelian theories (Mendie & Udofia, 2020; Cooper, 1985), as well as those by Martha Nussbaum (2011) and Richard Kraut (2007).

illustrated in the case above. At best, it might say nothing about the harm that Jack could do to Bob and other members. At worst, it might even encourage Jack to perpetuate such harm by encouraging him to focus solely on his benefit. Thus, for a theory of the digital good life to ensure the well-being of others, it must take into serious account the interconnectedness or relationality of current human existence, which is brought about in part by digital technologies.

However, I submit that the three dominant theories of well-being — namely, the hedonistic, desire-based, and objective list theories — might not be promising candidates for this kind of theory because they are susceptible to the aforesaid individualistic bias. The terms "well-being" and "good life" are interchangeable in the dominant philosophical literature (Kim, 2020). As such, even though these three theories are about "well-being," they can also be considered theories of the good life, and by extension, the digital good life.

The potential for individualistic bias is evident in the nature of desire-based and hedonistic theories. Desire-based theories get their name from the fact that they conceive of the good life or well-being in terms of what the individual agent desires. Hedonistic theories, in turn, base living well on the attainment of what is pleasurable for the individual. These reasons do not necessarily preclude the desire-based and hedonistic theories from considering the needs of others; it is conceivable that considering the needs of others is what one desires or takes pleasure in. However, they arguably promote a mainly self-regarding orientation, as these theories essentially tell them that whatever goods they should pursue in life are those that they are simply inclined to. This self-regarding orientation might then slide into an excessive focus on the self to the neglect of others – thus we return to the dilemma of Jack and Bob. On the other hand, individualistic bias may not be as evident in objective list theories, as they do not rely on individual preferences. They can posit "objective goods" which are good independently of the individual's dispositions.

Nevertheless, all three theories are susceptible to individualistic bias due to the basic structure they share. That is, all these theories consider the individual as the subject of flourishing.[7] To see why this can lead to individualistic bias, let me borrow a critique of Aristotelian virtue ethics by Thomas Hurka. Hurka claims that Aristotle's virtue theory suffers from explanatory egoism (2013). This means that the Aristotelian virtues only have value insofar as they contribute to that individual's flourishing. An

[7] See, for instance, the following literature, which discusses the different theories of well-being; all assume, implicit or explicitly, that the individual human is the subject of well-being: Parfit (1984), Bradley (2015), Haybron (2008), Kim (2020), and Brey (2012).

implication of this is an egoistic motivation in which the aspirant of virtue commits virtuous acts only for the sake of their own flourishing and not, say, due to the independent value of (promoting) other people's well-being. The theories of well-being we have discussed generally share the same structure as Aristotle's virtue ethics in that all the goods these theories list mainly have value only inasmuch as they contribute to the well-being of the individual who subscribes to them. Thus, it is also possible that these theories might lead to an egoistic – or individualistic – outlook, wherein the same individual mainly focuses on their own well-being while not giving due consideration to the well-being of others.[8] Moreover, since the focus of these theories is the individual, they are at risk of conceptually isolating the individual and portraying an atomistic view of human existence, which undervalues the importance of relationships and community in achieving the good life.

To clarify, I am not saying that focusing on the individual is *wrong*. Instead, it may be *insufficient* in capturing the relational or communal aspects of human living and flourishing, which seem deserving of more attention when considering how to live well in light of the various digital technologies that have increased human interconnectedness and the likelihood of people impacting each other's prospects for flourishing.

Considering this, what is needed is a philosophical approach to the issue of the digital good life that can counterbalance the potential individualistic bias of the abovementioned theories. Stated differently, this approach would consider the good life as supervening not only on the human individual but also on the network of relations in which the individual is situated. This might more accurately reflect the increasingly intertwined nature of human existence and flourishing made possible by the prevalence of digital technologies. I believe that such an approach can be inspired by Confucian philosophy, which I will demonstrate in Section 1.4. Before that,

---

[8] Another framework worth mentioning is the popular METUX (Motivation, Engagement, and Thriving in User Experience) model (Peters et al., 2018). This is because the framework is specifically designed to assess well-being in the context of digital technology use. Although this model includes "relations" as a component of well-being, it still considers the individual as the basic unit of well-being. Moreover, it only counts as "relations" interpersonal interactions, such as those in social media. It does not consider broader relations, such as one's relation to one's socio-political environment, which arguably also affects one's chances of living well. The METUX model does include "society" as a sphere of relation involved in digital well-being; however, it does not elaborate on this sphere, as its primary focus is "individual user experience [of digital technologies]" (Peters et al., 2018, p. 6). Considering these, the METUX model is also in danger of having an individualistic bias like the abovementioned theories of well-being.

however, let me discuss the second reason why I choose to use Confucian philosophy for addressing the question of the digital good life.

1.3. On the Importance of Including Non-Western Philosophies in Philosophical Discourse

As mentioned, the three prevalent theories of well-being that are likely to be used to address the question of the digital good life are the hedonistic, desire-based, and objective-list theories. These, however, predominantly originate from Western philosophers or philosophical traditions.[9] There are three reasons why this might be problematic. First, producers and consumers of digital products include non-Western peoples whose cultural beliefs may not easily align with Western philosophical ideas (cf. Wong & Wang, 2021, pp. 3-5; Dennis & Clancy, 2022). Thus, a conception of the digital good life wholly based on Western philosophies might not be easily understood by these people. Second, the exclusion of non-Western philosophies might prevent or delay the discovery of ideas that could improve the conception and attainment of the digital good life. This is a corollary of a more general notion that non-Western philosophies have ideas that might be absent or have been neglected by Western philosophy (cf. Wong & Wang, 2021, p. 2; Van Norden, 2017, pp. 5-7; Kim, 2020, p.21). Consequently, exploring non-Western philosophy can yield novel contributions and solutions to the issues that Western philosophy faces. For example, contemporary (Western) virtue ethics draws heavily on Aristotle's philosophy. However, Aristotle's understanding of virtue has a difficulty (Van Norden, 2017, pp. 65-66). Aristotle considers virtuous acts to be true when they stem from a genuine desire for virtue. At the same time, he considers human nature as ethically neutral. The difficulty lies in imagining how an ethically neutral nature can acquire genuinely ethical inclinations, such as the desire for virtue for its own sake. On the other hand, the Confucian philosopher Mencius asserts that human nature is not inherently ethically neutral. Rather, it possesses inherent ethical inclinations that can become virtues if cultivated properly. Thus, Mencius' idea of an ethically good human nature can serve as a resource for addressing the difficulties found in Aristotelian virtue theory. The third reason for including non-Western philosophies in the discourse of the digital good life is that it is simply a matter of justice. The general field of (academic) philosophy suffers

[9] Hedonist theories have their origins in utilitarian philosophers Jeremy Bentham and John Stuart Mill, Desire-based theories have their origins in the welfare tradition and objective list theories arguably originates with, and is most influenced by, Aristotle (Brey, 2015).

from a hegemony of Anglo-European philosophy that marginalizes non-Western philosophical traditions (Kirloskar-Steinbach & Kalmanson, 2021; Smith, 2015; Van Norden, 2017). Thus, it is only right that non-Western philosophies be included in contemporary conversations on issues that require a philosophical approach, particularly if such issues are of a global scale, such as that of the digital good life.

The reasons above motivate the decision to draw inspiration from Confucian philosophy to approach the question of the digital good life. This research aims to make a modest contribution to the inclusion of non-Western philosophies in contemporary philosophical discourse.

However, beyond the notion that prevailing theories of well-being are Western, one could also examine the literature on the digital good life and well-being, and ask if it overlooks non-Western philosophies. In general, non-Western philosophy appears to be underrepresented. For instance, consider *The Good Life in a Technological Age* (2012). This philosophical anthology employs a multidisciplinary approach in exploring how the good life can be achieved in relation to different kinds of technologies, such as "information technologies, medical and agricultural technologies, and consumer technologies" (Brey et al., 2013, p.4). Though its pluralistic approach is admirably broad, it does not utilize non-Western (philosophical) approaches. A similar observation can be made for a more recent work, *Ethics of Digital Well-being* (2020), which is also a predominantly philosophical anthology with a multidisciplinary approach but focuses specifically on digital technologies. Furthermore, a thematic review of the digital well-being literature by Burr et al. (2020) categorizes the literature into several domains and themes. However, none of these include or involve non-Western theories of digital well-being or the good life.

Nevertheless, there are works that apply non-Western philosophies to the issue of digital well-being. It is worth mentioning them here to give readers some idea of the presence of non-Western philosophy in the philosophical discourse about digital well-being, as well as to distinguish my research from these works. I remind my readers that the terms "good life" and "well-being" are often interchangeable in the literature. Thus, even if these works explicitly discuss digital well-being, what they intend by the term "digital well-being" can also include what I consider to be the digital good life. Because of this, I will not distinguish my research from these works simply because my research is ostensibly about the digital good life and theirs is on digital well-being. Instead, I shall distinguish my research from theirs based on the subject matter or topic.

One is Deepak Chopra's *Digital Dharma: How AI can Elevate Spiritual Intelligence and Personal Well being* (2024). In this book, Chopra uses AI language models as support for attaining a version of well-being based on Indian philosophy. Chopra's book, however, is limited to only the use of AI language models; he does not discuss how other digital technologies can help one achieve well-being or the good life. Moreover, his idea of how AI language models can help one to live well is limited to one method: simply asking questions about the good life to the language model. Although there is nothing wrong with this, there are nevertheless several other ways that digital technologies can influence and promote the good life. This dissertation attempts to explore these by looking at a wider range of digital technologies, as will be shown in section 1.4.

There are a few other works that use non-Western philosophies to promote digital well-being. They are as follows: Sarah Mattice (2015), who tries, with broad strokes, to apply Confucian, Daoist, and Buddhist ideas to social media for promoting well-being, Hui Jin and Edward Spence (2016) who use Daoist philosophy against internet addiction, and Matthew Dennis and Elenna Ziliotti (2022) who draw from Confucian perspective on human relationships to promote digital well-being in social media use. Furthermore, there are Phattharachai Uthaphan and Phramaha Phooriko (2024), and Somkhuan Namsithan and Patipun Junte (2024), who all apply Buddhist mindfulness to promote digital well-being for social media use. Finally, Leon Roets proposes a way to maintain digital well-being and health in relation to online communities through Ubuntu philosophy (2025). These works are all laudable for contributing to the inclusion of non-Western philosophy in the philosophical discussion of digital well-being. However, all of them focus on social media. Again, my dissertation will explore other digital technologies and their affordances, and how Confucian philosophy can be applied to them to promote the good life. Moreover, although my dissertation will also explore social media, it will nevertheless focus on a specific issue that the abovementioned works do not directly address, namely, political polarization.

So far, the works mentioned ostensibly discuss digital well-being. However, Shannon Vallor's book, *Technology and the Virtues* (2016), appears to focus primarily on what I consider to be the digital good life. Vallor draws inspiration from Aristotelian, Confucian, and Buddhist ethics to create a list of "technomoral virtues" that can help people shape their relationship with technologies in ways that promote or maintain the good life. Vallor shows how these virtues can be applied to specific types of technologies, namely, social media, surveillance technologies, war robots, and biomedical enhancement technologies. Vallor's knowledge and use of non-Western

philosophies are admirably comprehensive. Still, my dissertation will differ from her work in that its object of study will be more specific; instead of examining a type of technology in general, it will investigate particular issues or case studies about different digital technologies. Moreover, the dissertation shall be limited to recommending the cultivation of virtues; it will also propose design recommendations inspired by Confucian philosophy. Furthermore, this research can be considered more comprehensive in its use of the Confucian tradition than Vallor, as it will draw inspiration not only from Classical Confucian philosophy but also from Neo-Confucian philosophy.

Finally, the fields of Intercultural Information Ethics and Intercultural Digital Ethics are worth mentioning. The field of Intercultural Information Ethics, which emerged in the early 2000s from information ethics (Capurro, 2008), seeks to incorporate non-Western, indigenous, and Global South perspectives into discussions of privacy, autonomy, and responsibility, challenging Western dominance in the field. A subset of this, Intercultural Digital Ethics, focuses specifically on digital technologies and issues such as AI (Aggarwal, 2020). Intercultural Information Ethics and Intercultural Digital Ethics are relevant to the idea of the digital good life, as they explore ethical questions in technology while engaging with non-Western traditions, particularly Confucian philosophy. Scholars like Charles Ess have drawn on Confucian concepts, such as harmony, to support his *Pros Hen* approach (Ess, 2006, 2008), while Johnny Søraker has utilized Confucian ideas of relational existence to argue for the moral status of information and technology (Søraker, 2007). Pak Hang Wong has been a key figure in linking Confucian thought with Intercultural Information Ethics and Intercultural Digital Ethics (Wong, 2013, 2020, 2021) and advancing the concept of the "good life" as a framework for the ethics of technology (Wong, 2010). His work also challenges certain conceptual assumptions in ethics, such as the distinction between the good and the right (Wong, 2012). His anthology, co-edited with Tom Wang, titled *Harmonious Technology: A Confucian Ethics of Technology* (2021), also deserves special mention, as it attempts to apply Confucian ideas to practical issues in the ethics of technology, though it does not focus on the application of Confucian philosophy to the idea of the digital good life as this research does. Nevertheless, it can be considered a forerunner of and inspiration for this research.

To conclude, there has yet to be a comprehensive and sustained account of the normative insights that can be gained by employing a particular non-Western philosophical tradition to explore the concept of the digital good life and how to promote it. Again, Vallor's *Technology and the Virtues* comes closest to this. However, her

focus is on multiple philosophical traditions rather than on a deeper exploration of how one specific non-Western philosophy can contribute to the issue of living well with digital technologies. Thus, there still remains a gap concerning the contribution of non-Western philosophy to discussions of the digital good life. This dissertation is a modest attempt to fill that gap. This will be explained in more detail in the next section.

1.4. Drawing Inspiration from Confucian Philosophy to Address the Two Concerns or Gaps vis-à-vis the Question of the Digital Good Life

This dissertation aims to address the two concerns or gaps mentioned in sections 1.2 and 1.3 by drawing inspiration from the conceptual resources of Confucian philosophy. That is why the main research question of this dissertation is: **"What normative insights can a Confucian-inspired philosophical approach contribute to the pursuit of the digital good life?"** To reiterate, the two concerns or gaps are **(1)** the potential individualistic bias of Western philosophical theories of well-being and how this makes them ill-suited for being applied to the increasingly connected nature of human existence and flourishing afforded by digital technologies, and **(2)** the practical and ethical importance of including non-Western philosophy in current relevant philosophical discussions.

Answering the First Gap

To understand how drawing inspiration from Confucian philosophy can be a resource for addressing the potential individualistic bias in theories of well-being, an overview of the Confucian vision of the good life is necessary. Confucian philosophy considers human beings as irreducibly relational creatures (Ames, 2021). This means that their existence cannot be analyzed and evaluated without taking into account the relations in which they are part, particularly those with other human beings, such as family, friends, and fellow community members, including those in positions of authority. The Confucian good life consists of virtuously living out one's role in connection with these relations, which entails benefiting the people one has a role towards. To put it differently, the Confucian good life is an essentially moral one in which the central elements are moral virtues meant to perfect one's roles towards others, which entails benefiting them in some way. For instance, *xiao* (孝) or filial piety, fulfills one's role as a child to one's parents. This requires providing for them as well as showing them

deference and care (*Analects* 2.7; *Mengzi*, 5A1.5). Individuals fulfilling their roles toward each other virtuously ultimately yields the state of affairs called *he* (和) or harmony, in which people have mutually fulfilling relations without being forced into a uniformity (Confucius, 2017).

However, Confucian flourishing ultimately requires one to actively show goodness to others regardless of role. This can be seen in the highest, and the summation, of these virtues, namely, *ren* (仁), translated variously as "humaneness," "benevolence," and "goodness." Confucius did not precisely define *ren*; however, one salient characteristic of someone who possesses *ren* is caring for others as oneself (*Analects* 6.30, 12.2). This does not entail a merely "passive" disposition where one respects people as equals, should one encounter others. Instead, *ren* entails proactively seeking the good of others as part of one's own (pursuit of) flourishing. As Confucius says, "Desiring to take his stand, one who is Good [*ren*] helps others to take their stand; wanting to realize himself, he helps others to realize themselves" (*Analects* 6.30/ Confucius, 2003, p. 63).[10] Presumably, the flourishing person would possess the virtue of *ren*. Thus, the Confucian good life requires people to actively help others attain flourishing regardless of the latter's role.

These moral virtues, however, cannot be acquired in isolation. For instance, the virtue of *ren* is considered the outgrowth of the virtue of *xiao* (*Analects* 1.2). This is because one learns how to care for others as oneself (*ren*), first by exercising deep care towards one's parents, which is what *xiao* or filial piety consists of. Consequently, the existence of healthy family relationships is required for achieving *ren* and, by extension, the good life. Aside from family, teachers and moral exemplars are crucial for learning virtue. Indeed, Confucius served as both for his disciples in their pursuit of virtue (Olberding, 2012). Normative practices collectively known as *li* (禮) or "rituals" are also essential for virtue in that they provide a common way for both inculcating virtuous dispositions and expressing them (Ing, 2012). Finally, non-moral goods, such as basic material needs like food, clothing, and shelter, are also necessary for a good life. These might seem trivial, but Confucians emphasize their importance both in the sense that they are needed for living virtuously and that being virtuous includes helping others acquire these.[11]

---

10 Millenia later, the Confucian philosopher Dai Zhen would echo this, saying, "When a person fulfills his own life and then goes on to help others to fulfill their lives, this is humanity [ren]" (*Mengzi Ziyi Shuzheng*, Sec. 36/ Dai Z., 1990, p. 36).

11 E.g., *Mengzi* 1A7.20 – 1A7.24; *Mengzi Ziyi Shuzheng*, Sec. 10.

The above illustrates that the Confucian ideal of the good life is relational because it consists of the dynamic in which the individual and their social environment mutually benefit each other; the individual is expected to develop virtue in order to help their relations achieve flourishing, but these virtues can only be cultivated through the social relations and practices in which the individual is situated. In a certain sense, then, the subject of the good life from a Confucian view is the individual *and* the spheres of relation in which they are nestled. This is more explicitly suggested by the Confucian Classic, the *Great Learning* (*Da Xue* 大學). It says:

> The ancients, in wishing to manifest luminous virtue in the world, first brought good order to their states. In wishing to bring good order to their states, they first regulated their households. In wishing to regulate their households, they first cultivated themselves […] Cultivate the self and the household is regulated. Only after the household is regulated is the state well ordered. Only after the state is well ordered is the world at peace. (*Daxue* 4-5/ *Daxue & Zhongyong*, 2012, p. 135)

We can see in the passage that a virtuous person's influence should affect the four spheres of relation, namely, the self, the household, the state, and ultimately the world (see Figure 1).[12] This can be imagined as a process where the influence of the virtuous person expands towards ever-larger circles of relations. However, the passage can also be interpreted as saying that the virtuous person's virtue is *constituted* by its existence in those different spheres. This is because the manifestation of virtue only occurs after the ancients are able to positively affect the various spheres of relations. In other words, the subject of the virtuous life, which is essentially the good life in the Confucian view, *just is* the individual and these different spheres of relations in which she is situated.

[12] The passage in the *Great Learning* does not describe how the spheres of relation influence the individual. However, this idea can be gleaned from other Confucian literature which emphasizes the importance of having good friends and social environments in order to morally flourish (e.g., Xunzi 23. 380-390/ Xunzi, 2014, p. 257).

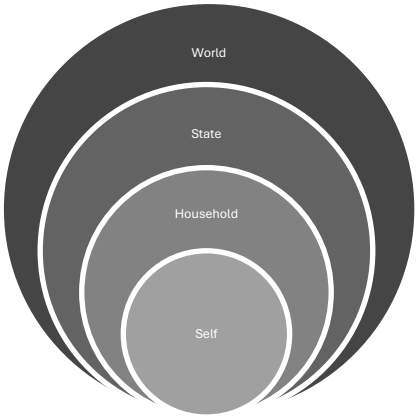


Figure 1

Confucian philosophy's vision of flourishing can thus inspire ideas that can counterbalance the potential individualistic bias of the dominant theories of well-being or the good life. One example is the idea that, in order to live well, one should care not only for oneself but for the people in one's sphere of relations because they are constitutive of one's good life and not just because they are conducive to one's individual flourishing. In the context of digital technology, this can mean that one should actively care for those with whom one is connected by virtue of these technologies, even if this does not necessarily or ostensibly benefit oneself. For instance, in the aforementioned case of an online job platform, a Confucian-inspired understanding of the good life would have no problem claiming that Jack should be considerate of fellow members, such as Bob, when applying for jobs. This is because these fellow members are part of Jack's sphere of relations, and their benefit is part of what it means for Jack to live a good life.[13] Ultimately, Confucian philosophy can be a promising resource for conceptualizing and advancing the digital good life, as it captures the relational nature of living well in the context of digital technologies.

[13] Pak-Hang Wong also uses (Neo) Confucian philosophy in arguing that a member of the online job platform has moral responsibility for fellow members with similar profiles (2021). He draws on the Neo-Confucian concept that there is no absolute separation between beings to motivate the idea that members of online job platforms should be more mindful of how their decisions can impact others. Although this reasoning is distinct from the one given in this dissertation (Pak uses the Neo-Confucian idea of oneness, whereas this dissertation uses the vision of human flourishing in the *Great Learning*), it nevertheless provides support that the Confucian philosophical tradition is an apt resource for formulating a less individualistic, more relational conception of human existence and flourishing.

Answering the Second Gap

Confucian philosophy is also a promising resource for addressing the second concern, namely, the need to include non-Western philosophies in the discourse on the digital good life. First, and most obviously, as a non-Western philosophy, the Confucian tradition is an apt choice for responding to the problem of Western philosophy's hegemony in the aforesaid discourse. Thus, using Confucian philosophy addresses one of the reasons for including non-Western philosophy in the discourse of the digital good life, namely, that such inclusion is merely a matter of justice against Western philosophy's dominance in the issue of the digital good life or well-being. Using Confucian philosophy can also address another reason for including non-Western philosophy in the discourse of the digital good life, namely, that wholly using Western philosophy for the said discourse might not be appropriate for non-Western peoples. Confucian philosophy is a salient part of the cultural heritage of several East Asian nations, such as China, Korea, and Japan, which themselves are major designers, producers, and consumers of digital technologies. A Confucian-inspired approach to the digital good life might yield insights that are more intelligible and amenable to the people of these regions than a Western philosophical approach. Finally, using Confucian philosophy addresses the last remaining reason as to why non-Western philosophies should be included in the digital good life discourse, namely, exploring non-Western philosophies might provide effective insights about the digital good life that would not be (easily) attained if thinkers limited themselves to Western philosophies. Confucian philosophy appears to be an especially promising example of this because, as shown above, its emphasis on the relational aspect of flourishing may produce insights or changes in perspective that the individualistic orientation of some Western philosophies of well-being/good life cannot (easily) attain.

1.5. The Methods for Answering the Main Research Question

Again, the main research question of this dissertation is **"What normative insights can a Confucian-inspired philosophical approach contribute to the pursuit of the digital good life?"** To clarify, the research will not use a historical Confucian conception of the good life as the basis for the idea of the digital good life. This is because the historical Confucian conception of flourishing might not entirely fit with contemporary social reality. Instead, this research will draw inspiration and insights from Confucian philosophy that can help people achieve the digital good life as described in section 1.1. To reiterate, this notion of the digital good life assumes the

prevailing general normative perspective that our global community largely holds today. As such, it assumes equality of human dignity and rights, as well as the chief importance given to people's attainment of well-being and harmonious or peaceful coexistence. Another feature of the digital good life is that it differs from the idea of well-being described above in that it can encompass moral goods, such as virtue. However, as I have argued for above, and as this dissertation will attempt to show, there is an important feature that dominant contemporary theories of well-being/good life miss and to which Confucianism can provide an important source of insight, namely, that the digital good life should encompass not just the individual but also the spheres of relations in which they are situated. Thus, I will add this feature to my conception of the digital good life. The more concrete way this is to be understood will be discussed shortly below.

Two further questions can be raised at this point. The first is, how exactly will the concepts from Confucian philosophy be used in this research? The second is, how will the main research question be specifically answered? Let me begin with the first question.

Drawing Inspiration from Confucian Philosophy

As has been mentioned several times, Confucian philosophy shall serve as "inspiration" for addressing the question of the digital good life. By "inspiration," I mean that I will not simply use concepts from Confucian philosophy as they were understood in particular historical periods or by other Confucian scholars. Instead, I will flexibly choose and/or modify elements of Confucian ideas to fit with the contemporary setting. This aligns with Confucian philosophy itself, as flexible adaptation to the context is a value highly prized by Confucians.[14] Moreover, adapting historical Confucian thought to contemporary times is an endeavor that several current scholars of Confucian philosophy have already done and are continuing to do, especially those working on the idea of Confucian ethics and democracy (Tan, 2003; Kim, 2023; Angle & Jin, 2025; Li, 2023).

[14] This can be seen in the Confucian virtue of "timeliness" (shi 時), which Confucius possessed, and which can be defined as "responding flexibly and appropriately to the situation with which one is confronted (Slingerland, 2003, p. 66). Confucian flexibility or adaptability will also figure in the way conceptual disruption is addressed in chapter 5 of the dissertation.

On Answering the Main Research Question

To answer the main research question, this dissertation will explore four case studies or issues where digital technologies influence their users in ways that potentially affect their chances of flourishing. This research will then draw inspiration from Confucian philosophy to offer solutions to these issues that can help users live well in relation to these technologies. The insights gained from applying this Confucian-inspired approach to these issues will serve to answer the main research question. These issues shall be picked based on a schema of the digital good life's scope, inspired by the *Great Learning*. Again, one insight that can be drawn from Confucian philosophy regarding the conceptualization of the digital good life is that the subject of such a life should not just be the individual, but also their spheres of relations. This schema concretely reflects this because it consists of four spheres of relations that, together, constitute the subject or scope of the digital good life. The four spheres are the spheres of the self, interpersonal relations, the political, and the world (see Figure 2 below). Each issue will belong to one of these spheres based on which sphere the digital technology of each issue affects the most.

The issues, along with their respective technologies and spheres of relations, are: "McMindfulness" for mindfulness apps, in the sphere of the self; toxic behavior for online multiplayer video games, in the sphere of the interpersonal; political polarization for social media, in the sphere of the political; and conceptual disruption for AI systems, in the sphere of the world. Table 1 below summarizes this.

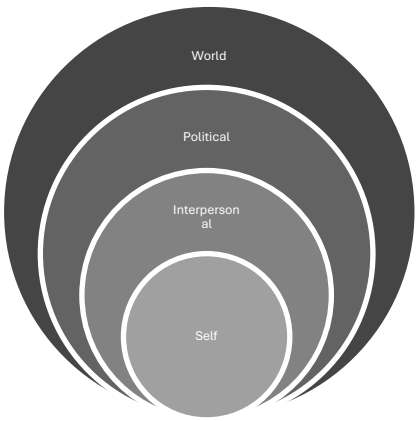


Figure 2

Spheres of Relations	Digital Technologies	Issues
Self	Mindfulness Apps	"McMindfulness"
Interpersonal	Online Multiplayer Games	Toxic Behavior
Political	Social Media	Political Polarization
World	AI Systems	Conceptual Disruption

Table 1

To reiterate, the research will draw inspiration from the conceptual resources of Confucian philosophy to provide insights that can help people live well in relation to these issues and the digital technologies associated with them. These insights can be classified into two. The first are those that apply to a particular issue/technology/sphere. I call these the *specific insights.* The second are those that apply to all issues/technologies/spheres, and can be identified by drawing the common ideas among the *specific insights*. I call these the *general insights*. Both of these insights will serve to answer the main research question.[15]

It might be asked why I modified the spheres of relations from those outlined in the *Great Learning*. In the *Great Learning*, the household (*jia* 家) is the sphere of relations just after the self because, traditionally, the most accessible relationships that Confucians had were those in one's family or clan due to physical proximity. Today, however, due to greater ease of mobility, different and more varied social organizations, and advanced modes of communication (including digital ones), relationships beyond one's clan can be just as accessible. Indeed, in contemporary usage, *jia*, as well as other Chinese familial terms, are used to refer not only to kin but also to friends and colleagues, that is, people with whom one might have personal relationships (Chen, 2019). In light of this, I have generalized the second sphere to encompass interpersonal relationships.

I have also changed the sphere of the state (*guo* 國) by renaming it into "political" to emphasize the fact that "state issues" and political issues are no longer rooted within the physical borders of the state. For instance, thanks to digital technologies like social media, people of the same state can argue over, and be affected by, the issues of their state even if they are not actually living there. Conversely, people can contribute to

15 As will be shown below, the *specific insights* also answer the subquestions.

the welfare of the state even while abroad due to the same technologies. However, due to the globalized nature of current society, what are considered political issues are no longer confined to the state. Rather, they can include phenomena that encompass multiple states, such as the relationship between nations, and political movements or concerns that several states have in common. Digital technologies, like social media, can track these international political issues and allow people to participate in the discourse about them. In this way, digital technologies also open up the possibility for people to affect these international issues. Because of these considerations, the political sphere I have formulated is meant to include even the aforementioned international political issues that go beyond the state's boundaries.

Regarding the fourth sphere of relation, namely, the "world," the scope is arguably broader than the third, in that it encompasses not only political issues but also all realms of people's interactions with their global environment. Both the digital technology and the issue assigned to this sphere of relation fit with this understanding of the world. Regarding digital technology, specifically AI systems, it aligns with the idea of the world, as AI systems have become global in reach, both in the sense that they permeate almost every aspect of human life and because their presence can be found worldwide. The issue of conceptual disruption, on the other hand, also fits with the idea of the world since concepts are one of the primary means by which human beings understand and navigate their environment (i.e., the world). Thus, conceptual disruption, as an issue that affects concepts, also affects people's fundamental perception of and interaction with the world.

1.6. Outline and Sub-Questions of the Dissertation

The four modified spheres discussed in the previous section will be the subject matter of chapters 2-5 of the dissertation, while chapter 6 will be the conclusion. Below, I shall give brief descriptions of these chapters. For the descriptions of chapters 2-5. I shall first identify the issue by articulating it in the form of a sub-question. The answer to this sub-question will help answer the main research question by providing the *specific insights* that a Confucian philosophical approach can contribute to advancing the digital good life in a particular sphere of relation. I will then explain the aforesaid issue, why it fits with the sphere of relation to which it is assigned, and how Confucian philosophy might address it. As for chapter 6, its main purpose is to summarize the *specific insights* of chapters 2-5 and to draw out the commonalities among them, thereby articulating the *general insights*. Both of these insights shall then provide the answer to

the dissertation's main research question, namely, **"What normative insights can a Confucian-inspired philosophical approach contribute to the pursuit of the digital good life?"**

In chapter 2, the sub-question to be answered is, **"How can conceptual resources from Confucian philosophy inform the efforts of designing mindfulness apps against 'McMindfulness'?"** The issue of "McMindfulness" aligns with the sphere of relation of the self because it affects users' perceptions of themselves and their roles within their social context. The phenomenon of "McMindfulness" is the condition in which Western mindfulness practices, including those in mindfulness apps, help people conform to unjust and onerous social structures instead of aiding them in evaluating and changing these structures. This endangers people's chances of flourishing by helping perpetuate a stressful and inequitable environment. "McMindfulness" is also contrary to Confucian ethics since the latter requires a person to enhance the quality of life of her context. Chapter 2 will thus suggest a Confucian-inspired design for a "socio-ethically aware" mindfulness app.

In chapter 3, the sub-question to be answered is, "**What normative insights can Confucian philosophy offer when it comes to mitigating toxic behavior in online multiplayer video games?"** This issue of toxic behavior in online multiplayer video games aligns with the sphere of interpersonal relations, as it affects the quality of interpersonal relations between video game users. Toxic behavior in these games is inimical to the mental and emotional health of the players who experience such behavior. From a Confucian perspective, toxic behavior is also undesirable since it diminishes interpersonal respect and propriety, which are chief Confucian goods. Thus, chapter 3 will suggest Confucian-inspired design features for video games that might mitigate toxic behaviors and promote respectfulness among players.

In chapter 4, the sub-question to be answered is, **"How can ideas inspired by the Confucian cognitive and ethical view of emotions contribute to reducing political polarization on social media?"** Needless to say, the issue of political polarization in social media fits with the sphere of the political. Political polarization can be problematic, as it hinders cooperative and harmonious social relationships (Ziliotti, 2022). Thus, chapter 4 will suggest ideas from Confucian philosophy that can inspire user strategies or digital design features that reduce political polarization and promote respect and sympathy instead. Moreover, considering that online polarization is significantly motivated by emotions, this

chapter will draw specifically from Confucian ideas about the cognitive and ethical nature of emotions.

In chapter 5, the sub-question to be answered is, **"How can Confucian views on concepts inform the efforts of engineering resilient concepts against conceptual disruption afforded by AI systems?"** The issue of this chapter is the possibility of conceptual disruption brought about by AI, which in turn can affect central concepts used in AI ethics and governance. This issue belongs to the sphere of the world because AI systems have become prevalent throughout the globe and affect almost all aspects of human life. Moreover, regulatory frameworks for AI have reached an international scope, such as the European Union's AI act. Furthermore, concepts are an essential means for our understanding of the world. Consequently, conceptual disruption affects our perception of, and interaction with, the world. AI systems have the potential to bring about conceptual disruption because it has profoundly challenged human practices or relations. For example, the concept of "friend" has been challenged because of the possibility of having AI as friends. Although the disruption of the concept "friend" (as with conceptual disruption in general) is not necessarily bad, it can have negative effects. For instance, it can leave one uncertain whether or not AI can indeed be friends with human beings. This might be detrimental to people who, due to their circumstances, have a need to form friendships with AI. Chapter 5 will thus explore how the resources of Confucian philosophy can help formulate a resilient concept, that is, concepts resistant to conceptual disruption. It will also utilize this Confucian-inspired resilient concept to address the issue of AI friendship by enabling people to consider AI as friends, provided it is conducive to their flourishing.

The dissertation will finish with chapter 6, the conclusion, which will answer the main research question: **"What normative insights can a Confucian-inspired philosophical approach contribute to the pursuit of the digital good life?"** The Confucian-inspired answers to the subquestions already provide the *specific insights* that contribute to answering the main research question. The conclusion will summarize these *specific insights* and, moreover, draw out the common ideas among them, thereby articulating the *general insights*. These two sets of insights, then, will constitute the answer to the main research question. The conclusion will also include additional ideas yielded by the research, but that do not directly answer the main research question, such as future areas of study that the limitations of the research suggest. Moreover, the conclusion will articulate the societal significance of the dissertation in terms of its contribution to the Ethics of Socially Disruptive Technology

(ESDiT) consortium, of which the dissertation is a part. The conclusion will end with a brief reflection on the entire research.

Overall, this dissertation aims to enhance the literature on the ethics of digital technology by examining issues from a Confucian perspective. It also contributes to discussions on the digital good life through a comprehensive Confucian lens. Furthermore, it aims to enhance the representation of non-Western philosophy in both the ethics of technology and contemporary philosophical discourse.

Interlude 1

It should be noted that chapters 2-5 are based on co-authored papers, all of which I am the first author. All of the co-authors have graciously given their permission for me to use these papers in my dissertation. More information on the authors and the publication status of the papers can be found on page iv. Considering the co-authored nature of chapters 2-5, the first-person pronouns in those chapters, that is, the pronouns used by the authors to refer to themselves, are in the plural form. However, the first-person pronoun shall return to the singular form in the Conclusion chapter since that part was written only by me.

Having said the above, the next chapter will explore the first digital technology to be discussed in this dissertation: mindfulness apps. Mindfulness apps belong to the sphere of the self because they are often designed for personal use and, as their title suggests, facilitate mindfulness practices that shape the self and one's understanding of it. The "self" that is formed through contemporary mindfulness apps, however, might be detrimental to the digital good life. This is explained by the critique of mindfulness practices known as "McMindfulness." This critique argues that mindfulness products perpetuate existing social inequalities and onerous lifestyles by focusing solely on individual self-improvement rather than addressing the broader social influences on well-being. Consequently, mindfulness apps shape one's perception of the self and others in a way that promotes unhealthy individualism. Confucian philosophy offers valuable insights that might be able to counteract McMindfulness by emphasizing the social context of human existence as essential to achieving the good life. Confucian thought emphasizes that social environments significantly impact an individual's ability to flourish and that enhancing the community is a crucial part of the individual's pursuit of flourishing. The chapter will demonstrate how these Confucian principles can inform the design and use of mindfulness apps. By introducing a more relational perspective of the self through mindfulness apps, Confucian philosophy can serve as a valuable resource for mitigating the individualistic tendencies of McMindfulness, thereby helping advance the digital good life in the sphere of the self.

Chapter 2. Counteracting Digital McMindfulness from a Neo-Confucian Perspective[16]

2.1. Introduction

Mindfulness apps are booming. They have an annual growth rate of 8.49% and are expected to reach 5.11 billion USD of global revenue within 2024. This is predicted to rise to 7 billion USD by 2028. The popularity of these apps has led to several studies on their effectiveness in dealing with mental health, which they claim to improve. Although their results are not definitive, these studies have provided evidence that mindfulness apps improve mental health by alleviating stress, depression, and anxiety and improving sleep quality and emotional regulation (Lahtinen et al., 2021; Huberty et al., 2021; Keng et al., 2022; Eisenstadt et al., 2021). The capacity of these apps to improve mental health supports the claim that mobile health apps can promote wellness or well-being.

However, not all evaluations of mindfulness apps are positive. Gill and Donaghue view them as a temporary but ineffective means of coping with the stressful environment brought about by neoliberal influence (2016). They are ineffective because they distract people from realizing that the root of their stress lies in systemic factors. Relatedly, Slunecko and Chlouba argue that mindfulness apps have been designed to subtly inculcate onerous neoliberal ideas into their users, the most pertinent of which is that success in current society lies in perpetually enhancing one's performance and productivity (through mindfulness) (2021).

These criticisms point to a broader phenomenon that has affected contemporary Western mindfulness, namely, 'McMindfulness.' McMindfulness refers to how Western mindfulness practices have been commodified and stripped of their original Buddhist religio-ethical foundations, thereby becoming a potential means for reinforcing neoliberal ends (Purser, 2019; Hyland, 2017; Forbes, 2019). Critics of McMindfulness argue that this is problematic since neoliberalism promotes excessive

[16] This chapter is based on a paper co-authored with Matthew Dennis, with the author of the dissertation as the first author. The manuscript is currently under external review for inclusion in the anthology *The Future of Digital Wellbeing*. The co-author has given his permission to use this paper for this dissertation.

individualism, competitiveness, profit acquisition, and inequitable social structures (Davies, 2014; Wilson, 2018; Forbes, 2019). Moreover, neoliberalism engenders a stress and anxiety-conducive lifestyle due to how it emphasizes constantly improving one's competitive and productive capacities lest one be left behind by others. Mindfulness practices and products that instantiate McMindfulness can thus be ultimately self-defeating since they might reinforce the onerous way of life that they claim to spare people from.

A possible, though partial, solution to McMindfulness is to construct new mindfulness practices that can counteract its effects. These new practices need a strong ethical and socio-political orientation, and a connection to critical inquiry. Although much has been written on how to design this 'reformed mindfulness' practice, no literature has discussed how this can be done in relation to mindfulness apps. This is surprising considering their ubiquity, as illustrated above.

In this paper, we aim to address this lacuna by utilizing the ideas of the Neo-Confucian philosopher Zhu Xi (1130–1200 C.E.). We draw inspiration from his ideas on mindfulness to suggest how today's mindfulness apps can be improved. Specifically, we intend to sketch a design for a course that can be included in mindfulness apps. We call this course 'Zhu Xi-inspired mindfulness' or ZIM, and it is intended to be our primary example of a suitably reformed mindfulness practice. We also suggest Confucian-inspired app features that can complement ZIM. ZIM and these features can counterbalance the influence of McMindfulness in current mindfulness apps.

Despite 'mindfulness' having Buddhist connotations, scholars have identified a similar concept in Zhu Xi's thought: jing (敬). However, unlike the kind of mindfulness present in mindfulness apps, jing, and its associated ideas, which we will discuss in section four, has a salient ethical and socio-political orientation. Indeed, Zhu Xi's ideas about mindfulness were formulated in his attempt to reconstruct the popular meditation practices of his time to be more ethical and other-regarding, in line with Confucianism's goals (Tiwald, 2016). This is similar to the present situation of combatting McMindfulness by reforming contemporary mindfulness practices to be more socio-ethically oriented. Because of these considerations, we believe that Zhu Xi's quiet sitting serves as a promising conceptual resource for precisely reforming mindfulness.

We do not claim that Zhu Xi's ideas are the only solution to McMindfulness. We recognize that various traditions of thought, such as Buddhism (Hyland 2017; Nielsen 2021; Purser, 2019) and Judeo-Christian religions (Forbes, 2022), have been drawn

upon to reform mindfulness. Our project can be seen as a contribution to this by including Confucianism in the set of traditions that can be sought inspiration from in combatting McMindfulness. Formulating a Confucian way of addressing McMindfulness might be helpful for people not inclined towards Buddhism or the Judeo-Christian religions. Furthermore, people who live in Confucian-influenced cultures could find ZIM appealing (Cf: Peng & Zhang, 2022).

We also do not think that ZIM serves as the 'silver bullet' or the foolproof means to remedy McMindfulness as it exists in apps. Instead, we consider it a modest yet hopefully distinctive contribution to the endeavor. In particular, we are aware that making mindfulness apps more socio-ethically oriented requires more than the effort of philosophers. Rather, it requires interdisciplinary collaboration from people with expertise, such as designers, computer scientists, and tech professionals. This is especially true when it comes to the design aspects of apps, such as UX (user experience) and UI (user interface) – aspects which this paper will mention briefly (in section 5) but not focus on. We hope, then, that this article can be a first step towards interdisciplinary collaboration.

The paper will proceed as follows: In section 2.2, we discuss "mindfulness" in traditional Buddhism and contemporary mindfulness apps, and how the latter could become "McMindfulness." Section 2.3 discusses a possible solution to McMindfulness: reforming mindfulness practices to foster socio-ethical engagement and critical inquiry. We also identify the necessary features of this reformed mindfulness. This leads to section 2.4, where we describe Zhu Xi's ideas on mindfulness that can concretize these features. These ideas are: "reverential attention" (i.e., *jing* 敬), the ethical program of the *Great Learning* (*Daxue* 大學), the "investigation of things" (*ge wu* 格物), and "quiet sitting" (*jingzuo* 靜坐). In section 2.5, we describe how the aforesaid concretization of the features can be done and how these "concretizations" can be combined in the form of the ZIM course in mindfulness apps. We also recommend app features that can complement the ZIM course. In section 2.6, we address possible objections to our ideas. We then conclude with a summary in section 2.7.

2.2. Mindfulness and McMindfulness

2.2.1. On "Mindfulness"

"Mindfulness" is a translation of the Pali word *sati*, an important concept in Buddhist soteriology (Gethin, 2011). It can be defined as a "lucid awareness of the phenomenal field […] that allows the object to stand forth with a vivid and distinct presence" (Bodhi, 2011, pp. 22, 27). This is not purely passive since it requires actively "sustaining attention on the object" (Bodhi, 2011, p. 30). Nevertheless, mindfulness involves receptivity because one should be detached and non-reactive toward the object of attention and anything that enters the mind (Analayo, 2006). For instance, if something agitates the mind, one should simply acknowledge its presence and analyze it objectively, noting the causes for its arising and ceasing. Mindfulness as detached, objective observation is a means Buddhists employ to attain their soteriological goals. For example, in the *Sattipathana Sutta*, mindfulness is used to scrutinize various phenomena dispassionately so that practitioners would realize their impermanence and undesirability, thereby decreasing their attachment to the world (Analayo, 2006). One could argue that mindfulness has no ethical orientation since it only involves observation. However, this is not the case since traditional mindfulness is inextricably linked with the Buddha's eightfold noble path, including (morally) right view and effort (Bodhi, 2011).

More than a millennium separates classical Buddhist mindfulness from its counterpart in mindfulness apps today. However, there is still resemblance between the two.[17] For instance, the *Headspace* website operationally defines mindfulness as

[17] It should be noted that this chapter does not claim that Buddhism is the exclusive source of contemporary mindfulness in the West. Mindfulness can be found in other religio-philosophical traditions, some of which are typically considered "Western," such as Stoicism (Sellars, 2018), Christianity (Trammel, 2017; Sun, 2014), and Islam (Kamarulbahri et al., 2024). Mindfulness in these other traditions also has contemporary versions (Cavanna et al., 2023; Caroll, 2023; Aldbyani, 2025). This chapter focuses on Buddhist mindfulness because it serves to contextualize the critique of McMindfulness, which is largely a critique of contemporary, Buddhist-inspired mindfulness practices in the West. Moreover, even if there are contemporary Western mindfulness practices originating outside Buddhism, the Buddhist-inspired conception of mindfulness is arguably predominant in Western discourse. This can be seen, for instance, in several scholarly attempts to define "mindfulness" (e.g., Bishop et al., 2004; Chiesa, 2013; Kelly, 2022; Chems-Maarif et al., 2025; Nilsson & Kazemi, 2016). These all cite Buddhist tradition as a primary source for the contemporary (Western) understanding of mindfulness. Moreover, popular mindfulness apps such as Calm and Headspace draw their understanding of mindfulness from the Buddhist tradition (Zeitlin-Wu, 2023).

follows: "You're focused on the present, not ruminating about the past or worrying about the future. Your attitude is open, curious, and non-judgmental. That's mindfulness" (2024). Likewise, *Ten Percent Happier*'s website states: "Mindfulness is about relating in an open, non-judgmental way to whatever feelings or perceptions you're actually experiencing" (Michaelson, 2019). As can be seen, mindfulness in these apps still involves a certain sustained, detached (read: "nonjudgmental") observation of the present moment. The main difference, however, lies in the divorce of this kind of mindfulness from its religious roots. The *Calm* blog defines this sort of mindfulness as follows: "Mindfulness is a non-religious, non-dogmatic practice with roots that can be tied to Eastern contemplative practice" (2023). This "secular" mindfulness was partly brought about by Jon Kabat-Zinn, who, in 1979, introduced 'mindfulness-based stress reduction' or MBSR at the University of Massachusetts as a clinical method to address chronic stress and anxiety (Nielsen, 2021, p. 51). Kabat-Zinn derived MBSR from Zen and Vipassana meditation techniques (Kirmayer, 2015). However, he separated MBSR from Buddhism's religio-ethical framework, thereby secularizing and 'de-ethicizing' it, making it more palatable to the Western medical community (Nielsen, 2021, pp. 52-53).[18] Today, many mindfulness apps draw inspiration from MBSR, serving as a secular means of nourishing mental health.

2.2.2. On "McMindfulness"

It is helpful to analyze McMindfulness in terms of neoliberal governmentality, as McMindfulness critics have (Purser, 2019; Forbes, 2019; Ng, 2016; Walsh, 2018). However, we would need first to break down the formulation of "neoliberal governmentality."

"Governmentality" refers to how the dominant institutions of society influence people's conduct by shaping their subjectivity through various cultural means such as social norms, practices, arrangements, and knowledge systems (Ng, 2016; Wilson, 2018). The goal of governmentality is for people to voluntarily behave and self-regulate through internalizing these means, thereby conforming to the values and interests of the dominant institutions. These values and interests have varied throughout history, but perhaps the most prominent today are those representing neoliberalism.

---

[18] Kabat-Zinn defines mindfulness as "paying attention in a particular way: on purpose, in the present moment, and nonjudgmentally" (1994, p. 4). One can see how this inspires the definitions of mindfulness provided by *Headspace* and *Ten Percent Happier*.

"Neoliberalism," on the other hand, is a complex and multifaceted phenomenon. We do not intend to describe it comprehensively here, but only its aspects relevant to McMindfulness. In general, Neoliberalism "rejects the notion of society and the public good and endorses a competitive, individualist, self-oriented market approach as the way to meet all of one's needs" (Forbes, 2022, p. 13). Neoliberalism extends economic liberalism's logic of free-market competition to all facets of human life. In doing so, it construes humans mainly as individualistic consumers (of the goods offered by the market) and ignores the communal aspects of human life that are arguably essential to human flourishing and which should be impervious to the demands of the marketplace (Harvey, 2005). Instead, individuals are seen as self-interested agents who invariably seek to maximize personal benefit while being in constant competition with one another (Forbes, 2019; Bal & Dóci, 2018).

At the structural level, neoliberalism manifests in the elimination of welfare programs, deregulation of markets, and other government policies intended to maximize profit and increase competitiveness. At the cultural level, it manifests through the idea of "self as entrepreneur" – a self who, in order to achieve continuous fulfillment through consumption, has to individually engage in relentless self-improvement so that it can compete with others and remain marketable (Wilson, 2018; Dilts, 2011). This self alone is responsible for its well-being; if its well-being diminishes, it is not its environment that is to blame but itself. Neoliberalism has been shown to cause distinct forms of stress and anxiety (Wilson, 2018; Purser, 2019). People become anxious due to the ever-present threat of losing jobs without a safety net to fall back on. Stress is produced by employees' increasingly heavy workload and the worsening — and unjust — working conditions they have to put up with due to lack of regulation. The onerous demands of the idea of the self as entrepreneur and the extension of the logic of competition throughout all facets of life also engender stress.

This is where (secular) mindfulness comes in.[19] People use products such as mindfulness apps to eliminate stress and anxiety. Without denying their benefits, these apps can nevertheless serve as mere palliatives that worsen people's entrenchment in the neoliberal system (Purser, 2019; Forbes, 2019; Kristensen, 2017). The separation of these products from their Buddhist framework has left them without any substantial ethical foundation that could serve as a contrast to neoliberalism. Instead, mindfulness products present themselves as generally value-neutral methods for coping with

[19] From here on, when we refer to "mindfulness" in the paper, we mean secular mindfulness as described above.

contemporary life's burdensome and hectic nature. The non-judgmental nature of most mindfulness practices in these apps can further reinforce this value-neutral appearance. We say "appearance" since these products are, ultimately, not value-neutral. Instead, they are a means of reinforcing neoliberal goals and values. In other words, they are instances of neoliberal governmentality, which is what McMindfulness is.

For instance, by reducing stress and improving concentration, mindfulness can become a tool for making people adjust to the neoliberal game, which prizes constant productivity and efficiency. Indeed, mindfulness has been advertised and sold explicitly as such. For example, one mindfulness promotor, David Gelles, states in his book *Mindful Work* that "Mindfulness can be a source of employer value proposition and may in the long run provide organizations with a valuable tool to manage high burnout levels of employees" (Gelles, 2015, p. 97, as cited in Purser, 2019, p. 139). Mindfulness app companies such as *Calm* and *Headspace* fulfill Gelles' words by selling "corporate mindfulness products" tailored to help companies improve employee well-being and productivity.[20]

Mindfulness can also reinforce the neoliberal notion of self as entrepreneur. The design and rhetoric of mindfulness apps encourage the idea that eliminating stress through mindfulness is only an individual responsibility. For instance, *Headspace* contains several mindfulness exercises in its app that are packaged only and distinctly for individual use (Kołodziejska & Paliński, 2022). Moreover, the discourse found in *Headspace* consistently tells users they are responsible for their well-being, and that part of this responsibility is doing these mindfulness exercises (Kołodziejska & Paliński, 2022).

Other mindfulness apps, such as *Stop, Breath & Think*, also tacitly encourage the idea that stress is a private problem (Slunecko & Chouba, 2021). *Stop, Breath & Think* asks users to indicate their emotional and mental states before and after engaging in mindfulness. However, the app does not encourage users to ask why they are in these states. Consequently, these states are decontextualized, with only mindfulness identified as a determining factor. This makes it appear that mindfulness alone changes one's emotional state. Since mindfulness is a practice done individually by the user, the app also implies that the user, qua individual, is the only one that can

[20] These products are shown on the web pages of *Calm Business* (https://business.calm.com/) and *Headspace for Organizations* (https://organizations.headspace.com/) respectively.

influence—and be responsible for—their emotional and mental states.[21] Consequently, stress and other related difficulties are framed as having their root and solution in the individual. This individualization of stress is problematic, not only because it reinforces the notion of self as entrepreneur but also because it deflects scrutiny away from structural sources of stress, such as unjust working conditions, that could be motivated by neoliberal values (Forbes, 2019; Ng, 2016; O'Donnell, 2015). Thus, mindfulness also threatens to preserve neoliberal structures.[22]

The nonjudgmental attention that secular mindfulness consists of and valorizes could also inculcate passivity and lack of criticality among practitioners (Legett, 2022; Walsh, 2018). In the workplace, this could mean that recipients of corporate mindfulness products might interpret being nonjudgmental as toleration of unjust conditions (Titmuss, 2013). This is another way that mindfulness could preserve inequitable structures.

McMindfulness not only affects present mindfulness apps but might also affect future ones. This can be inferred from recent literature proposing novel frameworks for designing mindfulness apps. This literature continues to assume the nonjudgmental kind of mindfulness mentioned above. Thus, these frameworks' design recommendations do not involve features that guide users toward ethical notions.[23] This could make the apps produced by these frameworks susceptible to becoming instantiations of McMindfulness. For example, Zhu et al. formulate a comprehensive framework for designing mindfulness apps that considers mindfulness as a state where "one does not have to do anything but accept what is happening at the present moment," the achievement of which could help reduce "stress and other mental suffering" (2017, pp. 2686, 2687). The design features highlighted by this framework are thus those that contribute to achieving this amoral state, such as "bare attention

---

21 This idea is stated in an alternative but more explicit way by mindfulness champion David Gelles who says: 'Stress isn't something imposed on us. It's something we impose on ourselves' (Gelles, 2015, 97, as cited in Purser, 2019, p. 138).

22 An example of a company that tried to exploit this potential of mindfulness is Amazon. In 2021, Amazon launched 'AmaZen' which was a small booth put in the middle of one of their warehouses where employees could practice mindfulness and 'recharge their internal battery' (See: https://youtu.be/2X2yMDsQPDw?si=y-e5YEfXbPwn-emB). This was at a time when Amazon was known for its lack of decent working conditions, and harsh policies, such as preventing employees from using the washroom. Unsurprisingly, AmaZen received severe backlash for illustrating precisely the company's attempt to use mindfulness as a diversion from its ethically problematic structure. For more on this, see Carvalho & Grácio (2022).

23 See Lukoff et al., (2020), Niksirat et al., and (2019); Zhu et al., (2017).

to the present moment" and "nonjudgmental acceptance" (Zhu et al., 2017, p. 2686). In contrast, no design features are suggested to introduce users to ethical ideas.

In summary, due to their lack of substantive ethical foundations, mindfulness products such as mindfulness apps are in danger of becoming a form of neoliberal governmentality, that is, "McMindfulness." McMindfulness shapes individuals' subjectivities and conduct according to neoliberal principles. Through McMindfulness, users voluntarily regulate themselves to meet the demands of neoliberalism, such as being constantly competitive and productive. Moreover, by leading people to internalize ideas like the self as entrepreneur, and by valorizing nonjudgmental attention, McMindfulness helps preserve the inequitable neoliberal status quo by diverting critique away from its possibly harmful and disenfranchising structures.

2.3. Reforming Mindfulness

Despite the problem of McMindfulness, critics still see the potential of mindfulness for forming peoples' subjectivities in ways that could emancipate them from neoliberal influence (Purser, 2019; Forbes, 2019; Leggett, 2022). For instance, by helping one's mind to perceive more lucidly, mindfulness practices could help one recognize hitherto overlooked aspects of themselves and their environment. People can use mindfulness to become cognisant of subtle neoliberal or inequitable elements that might detrimentally influence their lives, allowing them to act against these (cf. Ng, 2016). Mindfulness can also help people contemplate and adopt values or forms of life contrary to neoliberalism (Walsh, 2018; Hyland, 2016). However, these possibilities can only be actualized by intentionally redesigning mindfulness practices or products.

Reforming mindfulness practices can, therefore, be one way to address McMindfulness. What exactly is the reformed mindfulness practice supposed to be? They should be the "opposite" of McMindfulness (cf. Purser, 2019; Forbes, 2016, 2022; Hyland, 2017; Leggett, 2022). Whereas McMindfulness helps promote neoliberal values such as individualistic consumerism and individualizing structural problems, the reformed mindfulness should be the contrary. We can define the reformed mindfulness as whatever mindfulness practice helps people recognize the intrinsic value of human community and the socio-political environment for their well-being. This, in turn, can lead them to reject attitudes that can harm such community and environment, like the individualistic consumerism that neoliberalism

espouses. It can also lead them to examine and improve their social context to be conducive to the well-being of their community.

From this definition, we can extrapolate the necessary features of the reformed mindfulness practice. One is an ethical orientation emphasizing the socio-political aspects of life as part of human flourishing (Purser, 2019; Walsh, 2018; Cannon, 2016). This orientation should inform practitioners that their socio-political context affects their lives and that positively affecting it is part of their well-being. This orientation should also promote values that lead practitioners to improve this context, such as compassion and empathy for those who suffer, particularly from inequitable structures (Purser, 2019; Forbes, 2019; Walsh, 2018). This orientation and its values could help undercut the neoliberal notion of self as entrepreneur, which promotes the idea that individuals should only busy themselves with personal fulfillment through consumption, and that social factors are irrelevant to living well. It can also counter the ability of McMindfulness to individualize social problems since it will help practitioners realize that their environment plays a role in their well-being.

Another necessary feature is a connection to critical inquiry (Ng, 2016; Walsh, 2016; Forbes, 2019). By this, we mean a practice of methodically analyzing and evaluating the ethically salient elements[24] of one's socio-political context in order to identify and act upon the roots of structural injustices and suffering. This can help practitioners become aware of the detrimental effects of neoliberalism and other inequitable forces. It can also help practitioners discover occasions for working for positive social change that ultimately redounds to their well-being.

Finally, the reformed mindfulness practice must motivate social engagement (Somers, 2022; Legett, 2022). This feature is essential if it is to bring about tangible social improvement. This feature is merely a corollary of the previous two; if one values contributing positively to one's social context and is made aware of ways to do so through critical inquiry, the next logical step is to take action.

In the next section, we shall discuss how Zhu Xi's ideas involving mindfulness can inspire ways of concretizing these necessary features.

[24] By "ethically salient elements," we mean those elements that affect peoples' prospects of living well or having a good life.

2.4. Zhu Xi's Ideas about Mindfulness

Zhu Xi has four ideas about mindfulness that can concretize the abovementioned necessary features. These ideas are "mindfulness/reverential attention" (*jing* 敬), the ethical program of the *Great Learning* (*Daxue* 大學), the "investigation of things" (*gewu* 格物), and "quiet sitting" (*jingzuo* 靜坐).

Scholars have likened Zhu Xi's idea of *jing* (敬) to Buddhist mindfulness, even translating it *as* mindfulness.[25] To differentiate *jing* (敬), however, we will call it by another translation — "reverential attention."[26] Reverential attention requires single-minded, solicitous, concentration on an object. It is thus similar to Buddhist mindfulness because both entail careful, sustained attention towards a thing so that it might be lucidly present to oneself. However, unlike the mindfulness commonly found in apps, reverential attention is not "nonjudgmental" because performing it entails an evaluative posture. Specifically, it requires one "to concentrate on a task or object of inquiry with a certain kind of respect, which […] undercuts selfish intentions, including ones that are easy to overlook or misapprehend" (Angle & Tiwald, 2017, p. 148).[27] For example, if one is paying reverential attention to another person, one would avoid treating that person merely as a means for one's advantage or according to one's prejudices. In positive terms, one would treat that person with "high dignity" (de Bary, 1981). This connects to another feature of reverential attention, which is that it cannot be divorced from ethical action. As Zhu Xi says, "Reverential attention and rightness [*yi* 義] are just one thing […] Reverential attention is like having two eyes; opening one's eyes and seeing things is rightness" (*Zhuzi Yulei*, ch. 12 as cited in Ivanhoe, 2019, p. 32). The metaphor suggests that reverential attention should naturally lead to *and accompany* right action, just like how eyes lead to, but are still present in, the action of seeing. In other words, even when one is carrying out the ethical action that reverential attention inspires, one should still be in the state of reverential attention.

---

25 See de Bary (1981), Tan, (2019), and Ivanhoe, (2019). In Ivanhoe, jing is only compared to mindfulness but not translated as mindfulness.

26 This is the term used by Ivanhoe (2019) and Angle & Tiwald (2017).

27 This arguably means that reverential attention is inherently ethical since it is constituted by an ethical, or at least proto-ethical attitude, namely respect or reverence. This, in turn, possibly differentiates reverential attention not only from the mindfulness found in apps but also from traditional Buddhist mindfulness since the latter seems to acquire its ethical nature by only being attached to other (extrinsic) elements of Buddhist practice, such as right view and right intention (Bodhi 2011, p. 31).

Reverential attention can also be directed towards abstract things, such as ethical ideas. Paying reverential attention to these ideas should also lead people to accomplish the ethical demands that accompany them. For instance, paying reverential attention to virtuousness can, and should, spur Confucians to cultivate virtue. Relatedly, an essential set of objects for reverential attention is the ideals or goals of Confucian ethics (cf. Tan, 2019). Thus, to better understand reverential attention, we must look at the Confucian framework of ethical development that Zhu Xi followed.

This framework is found in the *Great Learning* (*Daxue* 大學), a short treatise considered to be a concise summary of the Confucian normative vision (Chan, 1963). The steps for cultivating moral virtue in the *Great Learning* is as follows:

> The ancients, in wishing to manifest luminous virtue in the world, first brought good order to their states. In wishing to bring good order to their states, they first regulated their households. In wishing to regulate their households, they first cultivated themselves. In wishing to cultivate themselves, they first rectified their minds. In wishing to rectify their minds, they first made their intentions cheng 誠 (true, genuine, sincere). In wishing to make their intentions cheng 誠, they first extended their knowledge to the limit. Extending knowledge to the limit lies in investigating things [*ge wu* 格物] […] From the Son of Heaven down to the common people, all without exception should take self-cultivation as the root. (*Daxue* 4-5/ *Daxue & Zhongyong*, 2012, p. 135)

By investigating things, the aspirant of virtue accumulates moral knowledge, which in turn helps them rectify themselves and form a genuine commitment to goodness. This commitment is then manifested by their efforts to order their social circles, such as the family, state, and even the world. From this process, one can see that Zhu Xi's framework for ethical development includes socio-political involvement. By being attached to this framework, reverential attention also attains a socio-political orientation.[28] Just how is reverential attention connected to this framework? Simply put, the entire process of manifesting virtue laid out by the *Great Learning* should be performed with reverential attention.[29] All the steps in the process should be accompanied with reverential attention directed not only towards the individual steps but to the ultimate goal of attaining virtue and benefiting the world. This will motivate

---

28 As will be shown shortly, Zhu Xi's other ideas on mindfulness also gain an ethical orientation through this.

29 As Zhu Xi says, "Reverential attention penetrates throughout the higher and lower stages of purposeful practice. Though one has reached the level of a sage, reverential attention still is essential" (Zhuzi Yulei, ch. 7. as cited in Ivanhoe, 2019, p.25).

aspirants to focus on and enact precisely this ultimate goal and the steps toward it (cf. Ivanhoe, 2019, p.10; Tiwald, 2016, pp. 144-145).

The first step in this process is the investigation of things. This entails careful study of and reflection on various objects and activities. Mostly, the things to study were morality-related books (Keenan, 2011). However, the objects of study could also be social and political relationships, daily affairs, and natural phenomena. Zhu Xi believed that through investigating things, aspirants would uncover "Pattern" (*li* 理) within them. Neo-Confucians believed "Pattern" to be the organizing principle of the cosmos. Although one Pattern pervades the universe, its instantiations exist in all things (Angle & Tiwald, 2017). Pattern provides the harmonious interconnection between parts of a thing and between things themselves (Angle, 2009). These interconnections are what produce and nourish life. More importantly, Pattern is *normative*—knowing it would give one insight into the proper way things should interact, including how one should morally conduct oneself toward others. In Zhu Xi's philosophy, the goal of the Confucian is to ultimately assist Pattern in its generation and nourishment of life by embodying morally good qualities, such as the chief virtue of humaneness (*ren* 仁).

To enhance the investigation of things, Zhu Xi believed that aspirants should practice quiet sitting. This was a meditation technique for clearing and calming the mind that Neo-Confucians such as Zhu Xi practiced. There appeared to be no agreed-upon method of quiet sitting – at a minimum, it required sitting in a fixed position and clearing the mind of distracting thoughts, inclinations, and emotions (cf. Ivanhoe, 2019, p. 145). However, Zhu Xi was concerned that quiet sitting, as practiced by some of his peers, had led to self-absorption and detachment from socio-ethical matters. Thus, he incorporated it into the program of cultivating virtue. One way he did this was by connecting it to the investigation of things.[30] The clear and undistracted mind quiet sitting afforded could allow aspirants to investigate things more effectively. Quiet sitting can also be an opportunity for paying reverential attention to moral ideas. After emptying their minds through quiet sitting, aspirants can more clearly pay reverential attention to "principles of right action" (Ching, 2000, p. 121), such as humaneness. This can help them appreciate such principles and motivate them to live them out. Finally, quiet sitting can also help cultivate reverential attention. This is because quiet sitting allows aspirants to practice reverential attention

[30] This also meant that quiet sitting should be done with the attitude of reverential attention. This ultimately helped orient quiet sitting towards ethical ends. See Tiwald (2016).

in a calm environment. Although reverential attention can be practiced amid daily activities, one would arguably learn to do it better in a setting without internal and external distractions.

In summary, Zhu Xi's four ideas concerning mindfulness are meant to operate synergistically to lead aspirants to virtue. Reverential attention – or mindfulness – is what motivates and permeates the process of virtue cultivation, as outlined in the *Great Learning*. To act virtuously, however, aspirants must attain moral knowledge through investigating things. Investigating things, in turn, can be more effectively done by clearing the mind through quiet sitting. Quiet sitting is also an opportunity for cultivating reverential attention and the motivation to live virtuously.

2.5. Formulating a "Zhu Xi-inspired mindfulness" or ZIM

Next, we will move to concretize the reformed mindfulness practice's necessary features, thereby formulating ZIM – "Zhu Xi-inspired mindfulness." Before proceeding, we wish to clarify what 'drawing inspiration' is. This means we will not entirely adopt Zhu Xi's ideas as they were historically understood. Instead, we will flexibly modify elements of these ideas to fit with the current circumstances and the purpose of reforming mindfulness while hopefully still retaining the core insight of these ideas.[31] This might seem arbitrary. However, it is almost unavoidable for those wishing to use ideas from the past to make alterations to them so that they might fit with the present context and thereby be efficacious. An example is Zhu Xi himself, whose interpretation of Confucian philosophy differed from the Classical Confucians due to the influence of, and need to contend with, the rival schools of his day (Ivanhoe, 2019).

ZIM consists of four components inspired by Zhu Xi's four ideas. These ideas are "reverential attention" (*jing* 敬), the ethical program of the *Great Learning* (*Daxue* 大學), the "investigation of things" (*ge wu* 格物), and "quiet sitting" (*jingzuo* 靜坐). The following sections discuss how these ideas can help concretize the necessary features of the reformed mindfulness practice.[32] Section 2.5.5, however, describes how all these "concretizations" come together as ZIM and how the latter can be translated as a course in mindfulness apps.

---

31 Whenever we substantially modify Zhu Xi's ideas, we shall explain in footnotes why we think our change retains the core insight of these ideas. See footnotes 36, 38, and 41.

32 In other words, how they can become the components of ZIM.

2.5.1. Reverential Attention as an "Ethical" Mindfulness and Motivator of Socio-ethical Engagement

Unlike the other components of ZIM, we directly use the concept of reverential attention instead of merely drawing inspiration from it. Following scholars of Confucian philosophy, we consider reverential attention as a kind of mindfulness. In other words, in the ZIM practice, reverential attention *is* mindfulness. Considering reverential attention as mindfulness can help concretize part of the first necessary feature of the reformed mindfulness practice, namely, having an ethical orientation. This is because, as implied in section 2.4, reverential attention has a basic ethical direction that points to respectfully and unselfishly treating the other.[33] However, this basic direction needs to be further informed to help people become ethical. Thus, an ethical framework needs to steer reverential attention toward the right (socio-ethical) goals; in turn, practicing reverential attention toward these goals can motivate practitioners to pursue them. This framework is the second component.

2.5.2. The Great Learning's Vision as Inspiration for the Reformed Mindfulness' Ethical Orientation

The *Great Learning*'s vision of the flourishing life consists of a virtuous person ordering the circles of relations surrounding them. This "ordering" begins with their closest relations, the family, and then expands to larger circles such as the state. This ordering generally entails helping the people in these circles achieve the means to manifest virtue.[34] From this vision, we can draw inspiration for four principles that together would form a basic ethical framework for the reformed mindfulness practice. This would provide the latter with its first necessary feature of having an ethical orientation emphasizing the socio-political aspects of life as part of human flourishing.[35]

The first principle is that one's socio-political context is necessary for flourishing. There are two ways to understand this. The first is that flourishing consists of ethically

---

[33] This is in contrast with the mindfulness found in apps which is paradigmatically nonjudgmental awareness.

[34] In his commentary on the *Great Learning*, Zhu Xi says, "Manifesting the original brightness of innate virtue in the world is to cause the people of the world all *to have the means* to manifest the original brightness of their own innate virtue" (*Daxue* Comm. 1/ *Daxue & Zhongyong*, 2012, p. 139).

[35] Although reverential attention might already have an inherent ethical orientation, the latter is not necessarily connected with valuing one's socio-political context. The framework drawn from the *Great Learning* makes that connection.

improving one's socio-political context by helping the people in it flourish.[36] This can be traced back to the *Great Learning*'s vision just mentioned. The second is that the ethical quality of one's socio-political context invariably influences one's chances of flourishing. This can also be drawn from the *Great Learning*, though not in the abovementioned part. For instance, it mentions that the ethical character of political authorities will affect their people's moral and material condition. Greedy rulers, for example, would impart their vice to their subjects. A state suffering this vice would then be depleted of the resources necessary for the well-being of its citizens.[37] This first principle suffices to give the reformed mindfulness the necessary feature of having an ethical orientation that emphasizes the socio-political aspects of life as part of flourishing. The other principles help in realizing the task of improving one's socio-political context.

The second principle states that improving one's socio-political context begins with one's immediate relations and then expands to further social circles. This can be drawn from the *Great Learning*'s description of the virtuous person influencing their

---

[36] It could be objected that this is not in line with historical Confucianism since it believed that people had different normative roles to play in society, and not all these roles directly involved improving the socio-political context. We have two responses to this. One is that in using Confucian philosophy as an inspiration for reforming mindfulness, we do not intend to strictly follow all of its historical features. Rather, as mentioned above, we aim to use Confucian philosophy flexibly so as to fit more with the needs of the current context. Our second response is that the idea that flourishing requires improving one's socio-political context can nevertheless be derived from Zhu Xi's interpretation of the *Great Learning*. According to Zhu Xi, exercising virtue requires helping people have the means to manifest their own virtue (See footnote 34). This, though, arguably entails that one be concerned with the social environment that people are in since, from the general Confucian perspective, this environment affects peoples' chances of attaining virtue. From these considerations, it seems that the flourishing person should be concerned with improving the quality of the socio-political context for the sake of helping others attain virtue.

[37] "If one household is ren [仁/ humane], the whole state is moved to ren. If one household is complaisant (rang 讓), the whole state is moved to complaisance (rang 讓). If one man (the ruler) is greedy and violent, the whole state acts in a disorderly way" (*Daxue*, comm. 9/ *Daxue & Zhongyong*, 2012, p.161). "There has never been a case of a superior (ruler) loving ren and inferiors not loving yi 義 [righteousness]. There has never been a case of [inferiors] loving yi 義 and affairs not reaching fulfillment. There has never been a case of the wealth stored in the storehouses not being his (i.e. the ruler's) wealth" (*Daxue* comm. 10.4/ *Daxue & Zhongyong*, 2012, p. 175). Zhu Xi explains the last sentence as the idea that a virtuous ruler, who inspires a virtuous citizenry, will not have wealth that is inappropriately used. Other sources in the Confucian tradition are also sensitive to how the socio-political environment affects people's chances of attaining the good life. For instance, Mencius claims that people without a constant livelihood are likely to fall into vice. Thus, a good government should ensure that they precisely have this constant livelihood (*Mengzi* 1A7.20-22/ Mencius, 2008).

immediate social circle first and then the remote but larger ones. Although the *Great Learning* identifies the closest social circle as the family, the second principle would recognize one's immediate relations as including neighbors, friends, and colleagues. This can help the purpose of the reformed mindfulness because it can lead people to examine the well-being of their co-workers and intimates, and whether these are affected by inequitable social arrangements.[38]

The third principle states that one needs to cultivate virtues to improve one's socio-political context. This can be derived from the *Great Learning*, which considers particular virtues necessary for ordering the family and state. Indeed, the virtues this third principle requires can come from the *Great Learning* itself, such as humaneness and righteousness.[39] The virtue of humaneness, in particular, has sub-virtues that are desirable for remedying McMindfulness, such as compassion, public-spiritedness, and sympathetic concern (*Mengzi*, 2A6.5, 7A4.3/ Mencius, 2008; *Daxue* comm. 10.3, *Daxue & Zhongyong*, 2012). By requiring the development of virtues such as humaneness, the third principle further fulfills the first necessary feature of the reformed mindfulness practice because the latter also requires the said practice to have an ethical orientation that promotes values like compassion and empathy.

Finally, the fourth principle states that improving one's socio-political context requires careful study of its ethically salient features. This can be drawn from the fact that, in the *Great Learning*, investigating things is the first step in the process of

---

[38] Admittedly the idea that one's neighbors, friends, and colleagues constitute one's immediate circle of relations is not in the *Great Learning*, which only considers family or kin (*jia* 家) as the constituent. We have made this modification to make the *Great Learning*'s schema fit more with contemporary times where, arguably, people generally have more opportunities (compared to pre-modern China) to interact and establish close relations with non-kin through their work, education, or location. Interestingly, in contemporary Chinese society, close friends and colleagues who are non-kin are nevertheless treated as an extension of one's family, as is seen from the fact that they are addressed in familial terms. See Chen (2019). Thus, there is some justification for our modification.

[39] According to the *Great Learning*, the virtuous ruler King Wen lived out the following virtues: "In being a ruler, he came to rest in loving kindness [or humaneness] (*ren* 仁); in being an official, he came to rest in respect (*jing* 敬); in being a son, he came to rest in filial piety (*xiao* 孝); in being a father, he came to rest in compassion (*ci* 慈); in his dealings with the people of the state, he came to rest in trustworthiness (*xin* 信)" (*Daxue*, comm. 3/ *Daxue & Zhongyong*, p. 143, 2012). The *Great Learning* particularly considers humaneness and righteousness as necessary for ordering the state: "There has never been a case of a superior (ruler) loving ren 仁 and inferiors not loving yi 義 [righteousness]. There has never been a case of [inferiors] loving yi 義 and affairs [of the state] not reaching fulfilment" (*Daxue*, comm. 10.4/ *Daxue & Zhongyong*, p. 175, 2012).

developing virtue and ordering one's society. This principle thus requires a method of critical inquiry. This is the third component of ZIM, which is discussed below.

2.5.3. The Investigation of Things as Inspiration for the Reformed Mindfulness' Critical Inquiry

Critical inquiry is not just a requirement of the fourth principle above. The reformed mindfulness practice also requires a connection to critical inquiry as its second necessary feature. We believe that the investigation of things can inspire a method for this critical inquiry.

As mentioned above, investigating things requires searching for the instantiations of Pattern. Pattern, in turn, is the life-producing and life-nourishing connections between things.[40] From this, we can derive a method of critical inquiry that aims to explore and evaluate the factors affecting people's quality of life or their prospects for flourishing. First, the method will lead people to explore their social context and look for those "patterns" that enrich their quality of life.[41] It will then lead them to reflect upon these patterns and use the insights they gained from such reflection to identify other areas in their lives where similar life-enriching patterns seem to be absent. Finally, practitioners will be prompted to use their insights above to find a way to "uncover" the life-enriching patterns in precisely those areas where they appear absent.[42] In other words, they will be led to think of ways to improve their quality of life and that of people they know, where such quality seems to be suffering.

In line with the ethical framework inspired by the *Great Learning*, this method can begin with oneself and then one's immediate social relations, such as family, friends, and colleagues. Practitioners can examine the status of their well-being and those they encounter daily. In particular, they could focus on people whose quality of life seems to be suffering (including their own) and try to discover what kind of life-enriching pattern is lacking so as to have caused this. After examining their immediate relations,

---

40 Li (理) or "Pattern" has been the subject of different interpretations during its long history in Confucian philosophy. For a brief introduction to this, see chapter 2.3 of Stephen Angle and Justin Tiwald's *Neo-Confucianism: A Philosophical Introduction* (2017).

41 Since the instantiations of Pattern are everywhere, they should also be found within social and political arrangements. For instance, Pattern exists in the relationship between father and sons, and rulers and ministers (*Zhuzi Yulei*, ch. 60, as cited in Tillman, 2019, p. 124).

42 We use the term 'uncover' here to remain faithful to the Neo-Confucian idea that humans do not invent cosmic Pattern. Rather, it is disclosed to them through practices such as, precisely, the investigation of things.

users can do the same kind of inquiry towards larger circles of relations such as one's community and polity.

At this point, one could ask: what exactly are the things that affect the quality of life or the prospects for flourishing? This might require a comprehensive account of flourishing, which we do not have space to provide in this paper. Moreover, a comprehensive account might not be ideal if the method of critical inquiry is to be used by different people with a plurality of beliefs. A more general account that aligns with people's intuitions might be the better option. An example of this can be derived from what Confucian philosophy generally values, such as material and psychological well-being, moral virtue, harmonious and meaningful social relationships, and a just and ordered polity (Kim, 2020; Confucius, 2003). These, then, are the things that the method of critical inquiry can track.

One might object that a general account of well-being contradicts the objective understanding of Pattern that Zhu Xi historically held. My response is that such an account of well-being can still be objective. Indeed, the Confucian values mentioned above, which can constitute a general account of well-being, are meant to be treated objectively, much like how Confucians have historically treated such values as objective components of an equally objective and particular vision of flourishing.

The objective nature of this general account of well-being can be further understood by comparing it with the notion of Pattern as Zhu Xi and other Neo-Confucians understood it. As mentioned above, Neo-Confucians generally believed that one Pattern pervades the universe. This one Pattern is the totality of all life-generating and nourishing arrangements. However, Pattern has specific instantiations, and there appears to be no absolute principle that dictates how these specific instantiations should be – except that they should precisely engender and nourish life. This does not make the specific instantiations of Pattern non-objective. Instead, it merely makes them particular to the context. Indeed, it is because of this context-specific nature that the instantiations of Pattern cannot be exhaustively codified and must mainly be discovered through investigating things. Something similar can apply when using a general account of flourishing. The account can give practitioners an idea of what generally and objectively contributes to life generation and nourishment. However, the particular 'Pattern' that should obtain in order for life to flourish in their context is something they can mainly know through the practice of inquiry itself. This does not render the particular Pattern subjective, as it can be considered an objective corollary of the general account in a specific situation.

An alternative or complement to using a general account is to make the method of critical inquiry intersubjective. People can perform the method together and, through reflection and dialogue, reach shared conclusions and insights. This resonates with Zhu Xi's idea that discussion with others can better facilitate the discovery of Pattern.[43] That is to say, coming up with ideas for improving quality of life can be made easier if people do it together. Moreover, making the method of critical inquiry a shared activity can help counter neoliberalism's emphasis on individualism.

2.5.4. Quiet Sitting as a Means to Support the Other Elements of ZIM

As mentioned previously, Zhu Xi believed that quiet sitting could enhance the investigation of things, help motivate the attainment of ethical goals, and cultivate reverential attention. This would also be its central role in ZIM. Since Zhu Xi claimed that there is no specific method for quiet sitting (Ivanhoe, 2019, p. 145) and that it only requires sitting in a fixed position while attempting to still and clear one's mind, ZIM can adopt such a general conception of the technique wholesale. Indeed, existing meditation techniques in mindfulness apps might be able to stand in for quiet sitting since these techniques generally involve taking a fixed position and stilling one's mind. Doing this could make ZIM fit more with the other content of mindfulness apps because ZIM would then share a common technique with them.

2.5.5. Bringing all the Components Together as ZIM in Mindfulness Apps

Now that we have discussed all four components, namely (1) reverential attention, (2) the ethical framework inspired by the *Great Learning*, (3) the method of critical inquiry inspired by the investigation of things, and (4) quiet sitting, we are ready to bring them together to give an account of ZIM. Taken together, ZIM is a mindfulness practice informed by an ethical framework inspired by the *Great Learning*. ZIM's purpose is to motivate people to attain the goals of this framework by practicing mindfulness cum reverential attention. This means that the entire practice is intended to be performed

[43] Zhu Xi talks about this in the context of reading classic moral literature. He says: "To come through reading, to an understanding of moral principle [i.e., Pattern] […] we mustn't engage in idle talk, but some parts of our reading pose problems while some parts are clear, so we have to discuss it. Those who don't discuss it are reading without dealing with the problems. Let's compare the essential meaning of a text with the different explanations of it to find out which explanations are correct" (Zhu Xi, 1990, pp. 153-154).

with reverential attention towards the aforesaid ideals. One of ZIM's goals is for users to carefully examine and evaluate their socio-political context's ethically salient features, thereby knowing how to improve it. To achieve this, ZIM makes its practitioners perform a method of critical inquiry inspired by the investigation of things. To do this inquiry effectively, practitioners are led to practice the meditation technique of quiet sitting. Through this, practitioners achieve the mental clarity and composure to improve critical inquiry. Quiet sitting also provides an opportunity for exercising reverential attention toward the goals of the framework, thus potentially motivating practitioners toward socio-ethical engagement. Finally, quiet sitting can cultivate reverential attention itself because it serves as an opportunity to practice it.

ZIM, however, is meant to be a course in a mindfulness app. Apps such as *Calm* and *Headspace* include several courses for teaching ideas or skills. These courses usually consist of a series of video/audio clips meant to be watched daily. ZIM can be such a course, arranged in the following manner: The first session could summarize what ZIM is, explaining its ethical framework, parts, and purpose. The subsequent three sessions could explain the major parts of ZIM, such as reverential attention, the method of critical inquiry, and quiet sitting. To engage users, these sessions can ask them to practice these elements after they have been explained. The last session can then guide how to perform the entire ZIM practice itself.

This practice can proceed as follows: The user is first asked by the voiceover of the video/audio clip to sit in a comfortable position, relax, and clear their thoughts. However, since the activity is ethically significant, it also tells the user to assume a serious attitude. In short, the user is asked to pay reverential attention. The voiceover then asks the user to breathe deeply a few times, after which it tells them to still their minds and further clear them of thoughts. After a few minutes of quiet, the voiceover tells the user to think of an ethical concept – perhaps humaneness. It gently leads the user to appreciate the concept, perhaps by making the user imagine how it would feel to be shown humaneness, and what they would feel when showing humaneness to others. The voiceover then more explicitly asks the user to reverently pay attention to this concept. After a few minutes, while still being cognizant of humaneness and its value, the voiceover leads the user to conduct the method of critical inquiry as stated in section 2.5.3. While this happens, the voiceover urges the user to try to benefit their circles of relations, reminding them that the virtue of humaneness entails this. The ZIM ends with the voiceover prompting the user to follow up on whatever virtuous

intention they have formed while doing the mindfulness practice. It also reminds users to practice reverential attention throughout the day.[44]

The following app features can also complement the ZIM course:

Writing and Saving One's Insights in the App

A common practice of Neo-Confucians like Zhu Xi was to write down insights about their process of ethical development (Keenan, 2011). These notes could aid them in remaining on the ethical path by helping them reflect on the task of moral improvement. They could also share these notes with their fellows to receive feedback on how to conduct such a task. Considering this, the app can have a functionality that allows users to type and save the reflections they gained from doing ZIM. As with the Neo-Confucians, these insights can help users advance in the ethical path ZIM introduces. This functionality would also be feasible since mindfulness apps like *Calm* already have similar features where users can precisely type and save their insights. Users could also share their insights through different means of communication.[45]

Reading Morally Significant Literature

The app could have a recommendation system that suggests literature to the user that might provide them with moral guidance and inspiration in the situation they face. The recommendation system can know what situation users face by looking at their notes.[46] For instance, if users write about the need to change what they perceive to be their onerous and unjust working conditions, perhaps the app could recommend writings of people who have advocated for better workers' rights.[47] This feature draws inspiration from Zhu Xi's idea that reading books, specifically the Confucian classics and related moral literature, is essential to cultivating virtue because Pattern can largely be found in the moral insights of this set of literature (Ng, 2019). Analogously, users who read literature connected to the moral challenges they face might discern

---

44 This reverential attention can be directed towards the people they meet and also to ethical ideas/goals of ZIM.

45 These different means of communication will be discussed further below, under the heading *Different Means of Communication for Users to Promote Group Activities*

46 This recommendation system will activate only with the permission of the user to ensure their privacy.

47 In line with the Confucian/Zhu Xi-inspired nature of the features, perhaps the four Confucian classics, i.e., the *Analects*, *Mencius*, *Great Learning*, and *Doctrine of the Mean*, can serve as default recommendations for morally insightful and inspiring literature. That is, this set of literature will already be in the app, free for the user to read, regardless of the recommendation system.

Pattern in this literature. That is to say, they draw ideas on how to morally improve their own situation from these writings. Moreover, the proper act of reading can develop virtues, such as carefulness and patience, that can help aspirants in other areas of moral development (Wong, 2025). Zhu Xi urged students to read the classical texts carefully because it is only through this that they will discern Pattern in the literature correctly. Careful reading means repeatedly going over the text, not only to understand its abstract meaning but also to grasp what the writers of the text (who are often moral exemplars such as the sages) experienced. This ability to carefully understand the text and sympathize with its writers can be transferred to the ZIM practice itself, particularly the method of critical inquiry. That is, just as users should carefully understand the readings suggested to them, they can also apply the same carefulness in examining their social surroundings – including people – for the presence, or lack thereof, of Pattern. Zhu Xi also believed that proper reading of texts can inculcate patience in readers because such patience is necessary for gaining a careful understanding of these texts. The patience developed through proper reading can be transferred or applied to users' treatment of their own ethical goals. As mentioned above, the ZIM practice is meant to urge users to improve the quality of life in their social setting. However, this goal might not be so easily achieved, thereby requiring patience on the part of the users.

Emulating Moral Exemplars

Related to reading moral literature is the idea of emulating moral exemplars. Zhu Xi, as well as the Confucian tradition in general, considered emulating moral exemplars an important component of moral development (Olberding, 2012; Leung, 2020). It is through this emulation that Confucians can better live ethically or, in Zhu Xi's terms, better know and enact Pattern in their lives (Cf. Wong, 2025). For Zhu Xi, reading the Confucian classics and related literature is the means by which Confucians can emulate moral exemplars because it is by going over these texts that Confucians can get to know the exemplars' lives, experiences, and teachings. Applying this to the mindfulness app, the literature that the app can recommend users to read can also include biographies of exemplary people who faced the same or similar challenges that they are currently experiencing. By reading about exemplars' lives and how they faced difficulties, users can gain moral insight and inspiration that could help them navigate and remain constant in their efforts to improve their ethical condition and that of their surroundings.

Recitation of Quotes or Passages

When reading literature, Zhu Xi emphasized the importance of reciting the text repeatedly instead of just scanning it with the eyes (Ng, 2019). Reciting the texts helps one understand, internalize, and remember the moral insight or message of these texts better, arguably because of how recitation engages the bodily and even affective faculties of the reader. In light of this, the mindfulness app can have a feature where users can select a quote or passage, either from their own notes or the literature they are reading, that they can choose to recite at different times of the day. The app can remind users to recite this passage through an alarm or notification. The alarm can also be accompanied by a pop-up that states the passage itself. Moreover, users can set the frequency of this notification to ensure it does not become a distraction.

Different Means of Communication for Users to Promote Group Activities

ZIM, particularly its method of critical inquiry, could be more fruitfully done with others. There can be app features that facilitate this by allowing users to locate and communicate with one another. This would not be unprecedented since mindfulness apps, like *Insight Timer*, already have such functionalities. To wit, *Insight Timer* has a page titled "Community" where users can do the following: join or form groups, search individual users and add them as friends (thereby allowing them to communicate via sms), and participate in live video conferences. Similar features can be utilized to make ZIM a shared experience. A possible way these features can be integrated into ZIM is this: Once the user finishes the ZIM course, the app can suggest fellow ZIM practitioners that the user can contact. The first potential practitioners that the app would suggest are the user's close relations, such as their family members and friends. The app can advise users to invite these people to take the ZIM course or to simply teach them about ZIM themselves. The app would suggest ZIM practitioners who are relatively near the user's location, for instance, those who live in the same city. This aligns with the ethical principle of ZIM, which states that improving one's social context begins with nearby relations.

Once users have formed (or joined) a group, the group could have its page in the app, as is the case for groups in *Insight Timer*. On this page, members can post/send messages to each other, schedule meetings, and conduct group sessions. By using the functionalities of the page, they can thus schedule to do ZIM together or just one of its parts, such as the method of critical inquiry. Perhaps ZIM could be done via live video conference to make the interaction between members more dynamic and personal. Ultimately, practicing ZIM as a group can give users the idea that

mindfulness and cultivating well-being by eliminating stress can be a communal affair instead of a private one. This can counter McMindfulness' reinforcement of neoliberal individualism and the individualization of stress.

Informing Users of Ethically Important Issues Happening in Their Context

The app can also give the user the option of being informed of the ethically relevant issues in their context.[48] To do this, the app will ask permission from the user to know the latter's location.[49] The app could then send users push notifications containing links to pertinent news from local news websites.[50] This feature can complement the method of critical inquiry since it helps users become knowledgeable of what is happening in their social environment. It can also help users become more concerned with their socio-political surroundings, which accords with the ethical vision of ZIM. This feature will only activate occasionally – perhaps thrice a day — to prevent users from being distracted by this feature. An alternative would be for the feature to activate only when there is significant news about one's local community or neighborhood that warrants concern. This would prevent the app from bothering users unnecessarily if there are no pressing matters in their situation. In case there are several important matters that the user should know, the times the feature activates would still be limited to an amount that does not distract users, e.g., perhaps all the news can be shown together only on a single occasion.

A Fitting UX (User Experience)

Although this paper primarily focuses on the ZIM course and the Confucian-inspired features, it is also worth remarking on the user experience (UX) we envision for the ZIM course and the mindfulness app as a whole. These are merely broad strokes and are not meant to be exhaustive or definitive. As mentioned above, we leave it to other experts (such as those in digital architecture) to come up with more comprehensive

---

48 The ethically relevant issues will be framed and determined in terms of the abovementioned general account of well-being that, in turn, can be constituted by Confucian values/goods (see section 2.5.3).

49 If the user is wary of giving away their specific location, there can be an option where they only need to identify the general place where they are located – perhaps their country.

50 By pertinent news, we mean those concerned with societal, economic, and/or political matters. One could ask, however, as to how the app would verify the trustworthiness of the news sites. We admittedly have no space to answer this thoroughly in this paper. One possible way to mitigate the danger of receiving fake news is for the app to present news about an issue from different sites, such as those belonging to different political spectrums. In this way, the user could come up with a holistic and unbiased view of the matter.

ideas for these design matters. Since the subject matter of the ZIM course and mindfulness app is ethical development and how to improve the ethical quality of one's social environment, the general user experience should convey the seriousness of such a topic. The colors, fonts, and other visual designs employed by the app should not be too gaudy. Instead, it should be more toned down and simple so as not to distract or give the impression of triviality. The sound effects should also be conducive to serious reflection and meditation. In this sense, it need not be too far from the sound effects employed by other mindfulness apps, such as chime or bell-like sounds fitting for meditation practices.

However, seriousness is not the only mood that the user experience should convey. Thoughtfulness and optimism should also be promoted. Thoughtfulness can encourage users to reflect on the things that the ZIM practice requires them to contemplate, such as ethical ideas and the ethical quality of their surroundings. Optimism can lead users to hope that the positive changes the ZIM practice urges them to enact can be fulfilled. An example of a component of user experience that can reflect seriousness, thoughtfulness, and optimism together is the voiceover that guides users in the ZIM practice. This voice should serve as a gentle yet firm (seriousness) guide for users, encouraging them to ponder on the ZIM practice's objects of reflection (thoughtfulness), as well as persevere in doing the practice, given the good fruits that it can yield (optimism).

2.6. Responses to Possible Objections

We want to tackle two related objections since we believe there are similar responses to them. The first is that including ZIM in mindfulness apps risks making users dependent on the app. This can go against ZIM's goal of encouraging people to pay attention to and engage with their social environment. The second is that apps engender neoliberal individuality by their very design. Since apps are made for personal use (and also personalize their features based on user preferences), they inevitably individualize the activities users do through them, even if these were originally meant to be done communally (cf. Gidaris, 2023).[51] This can heighten the sense of self as an entrepreneur since users view these activities as private practices for self-improvement rather than shared endeavors that require consideration of others. This, then, would go against ZIM's aims.

[51] ZIM itself can be an example of this. The ideas from which ZIM derived were originally practiced in a community of teachers and students.

We recognize these are serious objections, and our answers do not claim to be "fool-proof." Having said that, we would like to point to the idea of "multistability" in responding to both objections. Multistability is the notion that "any technology can be meaningful in different ways to different users" (Rosenberger & Verbeek, 2015, p. 25). That is to say, technologies can have multiple stable meanings or functions based on the context. This does not mean that a piece of technology can just have any consistent meaning since its materiality limits this possibility. Nevertheless, we believe that it is not outside the realm of possibility for apps to have a stable meaning besides the ones the objections assume. In other words, although apps can make users dependent and promote a sense of individualism, this possibility is not inevitable. A mindfulness app, in tandem with ZIM and its features, can have an alternative meaning or stability.

Perhaps an example of this alternative stability is to view the app, along with the ZIM course and its complementary features, as temporary "scaffolds" meant only to help users learn ZIM and the relevant capacities associated with it (cf. Lukoff et al. 2020, 1559).[52] They are temporary because they are not essential to ZIM and should be set aside if one has already mastered the practice. This can lead users away from dependence on the app/ZIM course. The app and/or ZIM course can have features that help establish this stability. For instance, throughout the ZIM course, there can be a constant reminder that users should eventually learn to practice ZIM independently without the app. Regarding doing the method of critical inquiry as a group, the ZIM course can suggest that it would be better for group members to eventually find other means of correspondence besides those provided by the app. Perhaps it can suggest that the group meets physically to form a greater sense of community. This, in turn, can counteract the possibility of the app reinforcing neoliberal individualism. Another feature that can bring home the idea of a "temporary scaffold" is if the ZIM course and its complementary features can only be used for a limited time before becoming unavailable. This is not unheard of since mindfulness apps, like *Calm*, also have features where full versions of themselves are freely available for a limited time before one needs to subscribe to access them.

Treating the ZIM course and its complementary features as temporary scaffolds can also address other objections. For instance, it could be objected that the feature informing users of essential issues in their context might prevent them from developing the capacity to be critically aware of this context. In other words, the ability to be

---

[52] A scaffold is "an environmental entity that supports, enhances, or regulates some behavior or capacity" (Steinert et al., 2022, p. 13).

critically aware of the context is "outsourced" to the feature. This, however, is less likely to happen if the feature is presented as a temporary scaffold meant only to help users develop their capacities instead of replacing them.

Aside from these objections, other important questions about ZIM should be addressed. For instance, if ZIM will be included in existing apps, how will it be advertised to the groups that make these apps, especially considering that the ZIM course urges its users to cease using it? We acknowledge that this is indeed a genuine concern. However, due to the space limitations of this paper, we cannot address this and other remaining concerns satisfactorily. Furthermore, as philosophers, we do not feel that we alone have the competencies to do so since further refining ZIM's design and advertising it to groups would require expertise that we do not possess. Thus, we reiterate that our article is better understood as suggesting a conceptual starting point for interdisciplinary collaboration rather than a complete plan.

2.7. Conclusion

In this paper, we show how Zhu Xi's ideas about mindfulness can help mitigate McMindfulness in the context of mindfulness apps. McMindfulness refers to how contemporary Western mindfulness, separated from its religio-ethical roots, perpetuates a neoliberal way of viewing the world. One way of combatting this is to reform mindfulness practices to have the following features: (1) an ethical orientation that considers the socio-political aspect of human life as part of flourishing, (2) a connection to critical inquiry, and (3) a way to motivate social engagement. Zhu Xi's ideas on mindfulness can inspire ways of concretizing these features. We combine these "concretizations" to design a "Zhu Xi inspired" mindfulness course, or "ZIM," for mindfulness apps. We also suggest app features that can complement ZIM. ZIM and these features can then equip users against McMindfulness. Together, these can counterbalance the influence of McMindfulness in mindfulness apps. We acknowledge, however, that there are essential concerns about implementing ZIM and that those with design and advertising expertise might better address these. Nevertheless, we hope that this paper contributes a modest "first step" towards transforming mindfulness apps from being forms of McMindfulness into means for an ethical and relational vision of human flourishing.

Interlude 2

The previous chapter discussed how Confucian philosophy can inform the design and use of mindfulness apps to counter McMindfulness, thereby contributing to the pursuit of the digital good life at the level of the self. The next chapter will shift focus to the interpersonal sphere, examining how Confucian philosophy might be able to contribute to mitigating toxic behaviors in online multiplayer video games, specifically in *League of Legends*. Confucian philosophy is a promising resource for this because it values interpersonal respect and harmony as constitutive of the good life. Online multiplayer games belong to the sphere of the interpersonal because, unlike mindfulness apps, they are designed for multiple users (or players) and, therefore, establish and mediate the relationships between them. These games, particularly *League of Legends*, are known for fostering toxic behavior, which can negatively impact players' well-being or prospects of achieving a good life. The forthcoming chapter will draw on Confucian philosophy to evaluate the affordances of *League of Legends* that may lead to toxic behavior, and formulate recommendations that have the potential to promote respect and harmony among players. In doing so, the chapter will illustrate how Confucian philosophy might contribute to the pursuit of the digital good life in the interpersonal sphere.

Chapter 3. Addressing Online Gaming Toxicity from a Confucian Perspective[53]

3.1. Introduction

Developing human relationships conducive to moral cultivation is central to Confucian ethics. In classical Confucianism, the individual self is a dynamic "nexus of relations" partly constituted by social relations (Gergen, 2009, p. 55, as cited in Thompson, 2017, p. 889). The social relationships in which the individual is situated shape her attitudes which, in turn, influence the individual's development and engagement in relations. This understanding of the self entails that attaining moral cultivation is possible only through harmonious human relationships at different social levels, from the interpersonal to the cosmic level.[54]

The Confucian emphasis on human relationships has recently prompted some contemporary scholars of Confucianism to examine technology-mediated human relationships. For these scholars, the extensive development and sustainment of digitally mediated interpersonal relationships raise critical questions for Confucian ethics: How can the Confucian goal of moral cultivation be pursued in a digital and "hyperconnected era"? (Floridi, 2015). Furthermore, is a Confucian-informed view of digital technologies possible?

A growing literature has emerged in recent years in response to the latter question, as scholars of Confucianism have begun analyzing disruptive new technologies, such as social media and robots. For example, in "Confucian Social Media: An Oxymoron?," Pak-Hang Wong (2013) proposes redesigning social media following Confucian values. Focusing on the Confucian understanding of the self, Matthew Dennis and Elena Ziliotti (2022) develop a Confucian-inspired account of social media's structure of human relations. In "AI and the Confucian Conception of the Human Person: Some Preliminary Reflections," Sorhoon Tan (2019) uses Confucian

---

53 This chapter is based on a paper co-authored with Elena Ziliotti, with the author of this dissertation as the first author. The paper is published in the *Journal of Confucian Philosophy and Culture* (https://doi.org/10.22916/jcpc.2022..38.131). The co-author has given her permission to use this paper for the dissertation.

54 In the *Great Learning* (*Daxue* 大學), ethical cultivation begins with the self and gradually expands to greater relational circles, including the family, the polity, and ultimately the world (*Daxue* 4 – 5/ *Daxue & Zhongyong*, 2012, p. 135, 4-7).

insights to pinpoint critical deficiencies in robot-human interactions. Robots are also the focus of Qin Zhu, Tom Williams, and Ruchen Wen's "Confucian robot ethics" (2019), which proposes an innovative Confucian account of morally competent robots.[55]

These innovative works advance debates in Confucian philosophy and philosophy of technology. However, no study on the ethical implications of online gaming from a Confucian perspective appears to exist.[56] Yet, online gaming has become an important part of contemporary human life. As it is estimated that three billion people, or 40 percent of the human population, play video games ("Global Video Game" 2020), sociologists and game study scholars consider video games to represent "an expression of life and culture in late modernity" (Muriel & Crawford, 2018, p. 2).

Notably, so-called "toxic" games are often the most popular. These online video games contribute to morally deleterious behaviors, involving "abusive communications directed towards other players [. . .] and disruptive gameplay that violates the rules and social norms of the game" (Beres et al., 2021, p. 1). For example, *League of Legends* (henceforth, *League*) is an online multiplayer game known for breeding toxicity (ESB Staff, 2021; Brinks, 2020; Bhatnagar, 2020). However, *League* is one of the most played PC games in the world, with an average of 180 million monthly players (Spezzy, 2022). The popularity of toxic video games, such as *League*, calls for more ethical scrutiny. It is unclear whether Confucian ideas can contribute to understanding the root causes of video games' toxicity and help formulate design requirements for redressing it.

This paper advances the understanding of these issues by focusing on *League*, one of the world's most popular toxic video games. We use Confucian intellectual resources to (a) diagnose the problems with such a game and (b) propose remedies to its toxicity. Following Dennis and Ziliotti (2022), we present a "Confucian-inspired design proposition" to curb *League*'s toxicity. This method draws inspiration from Confucian ethical ideals without imposing specific Confucian moral ideas on technological design (Dennis & Ziliotti, 2022, p. 1). Consequently, a

---

55 Other examples of noteworthy works in Confucian ethics of technology are Wong and Wang (2021), Wong (2019), Mattice (2015), and Wong (2020).

56 To our knowledge, the only reported study on the relation between Confucianism and video games is Ferril Irham Muzaki's empirical study on Confucian ideas on popular video games in Indonesia (Muzaki, 1990). Furthermore, Robert Page (2012, 2016) uses Confucian philosophy as an explanatory lens for understanding Chinese players' attitudes and behavior in two computer games. However, these works do not evaluate nor recommend changes to the games based on a Confucian ethical perspective.

Confucianinspired method avoids questions of feasibility and justifiability that arise around a claim for a Confucian techno-design for pluralist and non-Confucian contemporary societies (Dennis & Ziliotti, 2022, p. 12).

This paper claims that at least three reasons make *League*'s toxicity problematic from the perspective of Confucian ethics. Its toxic environment hinders the cultivation of humaneness (*ren* 仁), sympathetic concern or understanding (*shu* 恕), and harmony (*he* 和). However, Confucian conceptual resources can be used to formulate at least three remedies for *League*'s toxicity. First, inspired by ideas in the *Mencius*, we propose to remedy League's valorization of killing by redefining characters' defeats such that players' non-moral inclinations can be used to pursue morally acceptable ends. Second, inspired by the concept of *shu* 恕, we suggest that toxic behaviors between fellow teammates can be corrected if players could put themselves in others' place. This idea can be realized through a feature that invites players reported for malicious behavior to rewatch the highlights of the match in which they exhibited negative behavior. Third, the idea of a code of conduct constitutive of the notion of rituals (*li* 禮) can inspire the development of new scripted procedures to instill mutual respect among *League* players.

Section 3.2 opens by describing *League* and the causes of its toxicity. Section 3.3 analyzes League's toxicity in terms of Confucian ethics, while section 3.4 discusses how Confucian conceptual resources can inspire three design recommendations to mitigate *League*'s toxicity. Finally, section 3.5 summarizes the paper and suggests future research areas.

3.2. *League of Legends* and Its Toxicity

Before moving to *League* and its design elements, some terminological and methodological clarifications on the basic assumptions of our analysis of *League*'s design elements are in order. Our analysis draws from philosopher of technology Philip Brey's approach to technology. Like Brey, we consider technological products unable to produce consequences deterministically (Brey, 2018, p. 1). Instead, their probability of having specific effects partly depends on their specialized function and the social context in which they are placed by virtue of that function (Brey, 2018, p. 1). For this reason, we shall refer to the consequences brought about by *League*'s design

elements as "affordances" to signify that the effects of technological products are precisely only probable (Brey, 2018, p. 2).[57]

League is a multiplayer online battle arena game in which two teams of human players fight against each other to attain a particular objective. *League*'s primary game mode is "Summoner's Rift" (Kou, 2020, p. 82). A standard match in Summoner's Rift consists of two opposing teams of five members each. Both teams have bases on the opposite sides of the game mode's map (Kou, 2020, p. 82) and three lanes connect these bases. Built on the lanes are towers that teams should destroy to proceed to the enemy team's base. Between the lanes are areas known as "jungles" where monsters reside. Players can acquire advantages if they defeat any monsters. Both team bases have a "nexus" structure that creates "minions." These minions are continually produced, and they proceed to the lanes where they fight the minions of the other base. The victory condition of the match is the destruction of the opponent team's nexus. To achieve this, players can pick over a hundred "champions" to play as in their team (Riot Games 2022a). In a team, most champions fight in one of the lanes where they face enemy minions and champions, while one champion goes to the jungle to fight monsters. All champions begin at level one, where they can often use only one ability. A champion levels up when it kills minions, monsters, and other champions. As it levels up, it acquires more abilities and its traits[58] are enhanced. Concomitantly, a champion gains more money because the latter can be acquired by personally defeating minions, monsters, and other champions. This money can purchase items that enhance the champion's traits.

The most prominent Summoner's Rift subtype is Ranked Solo/Duo submode or "ranked games" (Kordyaka et al., 2019, p. 2487). A ranked game allows individual players to become part of a ranking system in which they ascend or descend like a ladder. The "rungs" of this ladder consist of nine tiers: Iron, Bronze, Silver, Gold, Platinum, Diamond, Master, Grand Master, and Challenger (Kou, 2020, p. 82). These tiers are each divided into four divisions. A player begins in "Iron IV" and, through winning matches, ascends to higher divisions (e.g., Iron III), although she can descend to lower divisions by losing or not playing for some time. Achieving a high

---

57 By "affordances," Brey means that a given technological product "may afford, enable, allow, induce, stimulate, cause, necessitate or require certain events or state-of-affairs" (Brey, 2018, p. 2). In this paper, we will usually use the verb form, i.e., "affords."

58 A champion's trait, also called a "statistic" or "stat" for short, "is a number indicating how well a champion can do a certain thing" ("Champion Statistic" 2021).

rank is considered prestigious and motivates players to engage consistently in ranked games (Kou, 2020, p. 87).

Ranked games are known to be more toxic than other game modes. Multiple design elements explain ranked games' higher toxicity.[59] One is the ranking system, which evaluates single players based on collective activities. This creates an overarching individual goal for players, namely to get a high rank, and motivates them to pursue it (Kou, 2020, p. 87). However, this goal can be achieved only through a nonindividual activity: winning matches with a team. The achievement of the latter goal depends heavily on each team member playing well (Vars 2021, 3:25-42), such that a player's teammates are partly responsible for her progress or regress in the ranks (Paul, 2018, p. 119). For instance, a player who performs well in a game can still lose that match and have his rank level drop if his teammates underperform during the game. This situation affords a hostile environment in which some players impose an overbearing demand on their teammates to play optimally and where they become furious at them for the slightest mistakes. As a *League* player explains:

My friend and I just started trying ranked. We got into a match and told our team that we were new to ranked so we would play extra carefully. However, when my friend died to the opponent ADC once, immediately one of our teammates said we suck, we are feeding,[60] and wanted to report us after the match. (Kou, 2020, p. 88)

The so-called "snowballing" phenomenon further increases pressure on teammates. Snowballing means that "all actions have repercussions for the duration of the game and come one after another in an exponential way" (Bonenfant et al. 2018, p. 143). Thus, even a single failure or defeat on the part of one team member can significantly raise the chances of the team losing the match. This increases the demand for optimal performance, the tension among players, and ultimately the likelihood of toxic behavior

How the ranking system affects players' perception of each other also affords toxic behavior. The ranking system can dispose players to see their fellow players as means to achieve a high rank. This instrumentalizing attitude diminishes the dignity of fellow players, making the one who has this attitude more prone to treat them disrespectfully. This is especially true when fellow players make mistakes; that is when the "'instruments' fail to perform as expected" (Kou, 2020, p. 90). Furthermore, players

---

59 Kate Grandprey-Shores et al. show that players of ranked games "were associated with higher toxicity indexes than players in normal matches" (2014, p. 1365).

60 "Feeding" refers to how champions make their enemies stronger by constantly losing to them. This is possible since defeating a champion results in monetary reward and experience.

can use ranks as the basis of their conduct towards one another (cf. Kou, 2020, p. 87). Often, the presence of ranks leads to displaying greater trust and esteem for people in higher ranks. However, it also promotes belittling players of lower ranks. This manifests in abusive communication. For instance, a *League* player attests that "Some players love calling Bronze [ranked] players monkeys or apes" (Kou, 2020, p. 87).

Finally, the ranking system can also afford toxic behavior because teammates may not have the same motivation to win the match. Teammates may be at different places in the rankings, and this can lead to friction, a lack of cooperation, and disrespectful behavior between team members. For example, one player may only require one more game to get a higher rank, but she can be anxious about winning the match. Meanwhile, the rest of the team may have a more carefree attitude because they have several games to win before they move up in rank (Bonenfant et al., 2018, p. 149).

3.3. Evaluating *League* and Its Toxicity from a Confucian Perspective

Having clarified the game features that are usually associated with *League*'s toxicity, it remains unclear why these toxic features are problematic from a Confucian ethical perspective. This section (3.3) answers this question by revealing the most troubling aspects of *League*'s toxicity from a Confucian standpoint. It argues that crucial design elements of *League*'s ranked games hinder the cultivation of fundamental Confucian attitudes: disdain for killing, sympathetic concern, openness and accommodation.

3.3.1. Disdain for Killing and Appreciation for Life

From a Confucian perspective, one problematic element in ranked games is their portrayal of killing. *League* construes killing as necessary for winning a match and even valorizes a player's ability to kill multiple enemy champions. In contrast, the ancient Confucian masters had a substantial appreciation for the value of human life: for instance, both Mencius and Xunzi considered the virtuous person as someone who would not take an innocent human life even if he would obtain the whole world through it.[61] Mencius believed that a ruler should execute a person only if his entire country agreed to it and after he had examined the person.[62] Furthermore, Confucius

---

61 *Mencius* 2A.2:24 as translated in (Mencius, 2008, pp. 42-43); *Xunzi* 8:75-80, as translated in (Xunzi, 2014, pp. 54-55). The citations of Xunzi are based on Hutton's translation (2014). The number before the colon is the chapter, and the number(s) after it are the line number in Hutton's work.

62 *Mencius* 1B.7:56, as translated in (Mencius, 2008, pp.25-26).

believed that during ideal times, when truly virtuous rulers reign, capital punishment would be dispensed with[63] and the unjust would be converted to goodness through the ruler's powerful moral example.[64] The Confucians' disdain for killing can also be seen in their dislike of war. Mencius abhorred those who encouraged war and considered them guilty of a "crime even death cannot atone for."[65]

League seems to trivialize human life by making killing a necessary means of winning the game. One might object that since killing is merely instrumental for winning, players will not give it much importance. However, this is precisely the problem for Confucians: accustoming players to value human lives instrumentally to pursue their personal ends. Furthermore, *League* does more than construe killing as a means to win the game: it valorizes it by broadcasting a player's kills through the in game voice announcer ("Announcer" 2020). These announcements are based on the number of consecutive kills: two kills in a row are called "double kill," three kills "triple kill," four kills "quadra kill," and five kills, which is the whole enemy team, "penta kill" ("Announcer" 2020). The higher the number of consecutive kills, the more excited and admiring the announcer's tone becomes. Although it is difficult to ascertain the specific effects of *League*'s valorization of killing on a person's disposition toward human life, what is apparent is that it does not align with the Confucian appreciation for life and disdain of killing. *League*'s valorization of killing might also be a matter of concern for those who do not hold a Confucian perspective. This is because violence in video games, which includes killing, is believed to have some correlation to aggressive and antisocial behavior (Anderson et al., 2010; Greitemeyer, 2022; Prescott, Sargent, & Hull, 2018).

3.3.2. Sympathetic Concern

From the perspective of Confucian ethics, *shu* (恕), or "sympathetic concern" (Tiwald, 2010a), is a distinctive attitude of the virtuous person (*junzi* 君子).[66] Confucius prizes shu as the "method" by which one attains the supreme virtue representing moral perfection (i.e., humaneness/*ren* 仁).[67] Yet, some aspects of the ranked game mode can hinder players' ability to cultivate sympathetic concern. Indeed, shu entails caring

63 *Analects* 13.11, as translated in (Confucius, 2003, pp. 144).

64 *Analects* 12.19, 15.5, as translated in (Confucius, 2008, pp. 134, 175).

65 *Mencius* 4A.14:2, as translated in (Mencius 2008, pp. 96).

66 *Analects* 15.24, as translated in (Confucius, 2003, p. 183).

67 *Analects* 6.3, as translated in (Confucius, 2003, p. 63).

for others by imaginatively putting oneself in their place (Chan 2000, 509-510). This creative exercise consists of two different but integrated acts (Tiwald, 2010a, p. 83). The first is imagining "the thoughts and feelings that [a person] might have under real or imagined circumstances" (Tiwald, 2010a, p. 83). The other is imagining oneself as that person in those circumstances and caring for her with the same care one would apply to oneself.

However, cultivating sympathetic concern is difficult in ranked games. The system of individual reward, coupled with the instrumentalizing attitude discussed in section 3.2, can promote self-centeredness. It can incentivize players to focus on achieving their personal ends while deterring them from cultivating a genuine interest in their teammates. Under these conditions, developing sympathetic concerns toward other players is challenging. The ranking system aggravates the situation: if one sees others as of lesser worth than oneself, there seems to be little motivation for one to put oneself in their place. To be clear, instrumental relationships are not the problem from a Confucian perspective. Confucianism does not entail that individuals should value humans non-instrumentally under all circumstances (Tan, 2019). The problem with the instrumentalizing attitude in *League* is that it can promote an inconsiderate and rude attitude towards other players, making it difficult for dispositions like sympathetic concern to develop.

Finally, another element of ranked games that challenges the development of sympathetic concern is the attention-demanding nature of matches.[68] Sympathetic concern requires using the imagination and one's reflective capacities, such as when trying to determine the implications of applying one's self-care to another. This might take up attention and time, but players may be unable to make space for it if they are seriously playing a match. Players need to pay constant attention to the game since several things can happen within a few seconds that could give them an advantage or a disadvantage. For instance, enemy champions could ambush and kill one's champion if one becomes idle, even momentarily. Coupled with the snowballing phenomenon mentioned in section 3.2, this can give the enemy team a substantial lead. Thus, there is a demand for constant and complete attention to the match that leaves little space for practicing sympathetic concern. Of course, it is conceivable that exercising sympathetic concern could take only a little time. However, those who are still beginners to the journey of virtue-development (which presumably is most people) would require a longer time to perform this.

---

[68] This is true even for non-ranked games.

3.3.3. Openness and Accommodation

The elements of *League*'s ranked games also hinder players' ability to cultivate openness and accommodation, which are requisite attitudes for attaining interpersonal harmony. Generally, harmony (*he* 和) is viewed in Confucianism as the unity of a plurality of things in which the differences of these things do not negate each other but complement or balance each other (Li, 2014, pp. 28-29). This is vividly expressed in the *Zuozhuan* 左傳 (Commentary of Zuo), where harmony is compared to cooking stew.[69] In the passage, the word "stew" (*geng* 羹)[70] refers to a dish that has ingredients with intense flavors (Li, 2014, p. 26). However, in a good stew, these varying flavors do not overpower each other; instead, they are balanced so each can express itself while allowing the different flavors to do the same. Similarly, harmony is about diverse elements coming together and achieving an agreement that allows each to express their distinctiveness in a way that forms an "organic whole" (Li, 2014, p. 26).

Harmony also entails openness to novelty because it is not required to conform to a predetermined and absolute standard (Li, 2014, pp. 20-21). Furthermore, harmony does not exist only in inanimate objects such as the ingredients of a stew. It can also exist between humans and the natural world (Li, 2014, p. 17), among humans themselves (interpersonal harmony), and within the human individual (intrapersonal harmony) (Lewis, 2020, p. 64). Interpersonal harmony entails "a community of mutual respect and consideration" (Lewis, 2020, p. 134) in which the members' uniqueness is recognized as enhancing the group's quality (Confucius, 2017, pp. 55-56). For this reason, interpersonal harmony requires an attitude of openness to difference or, precisely, uniqueness. Included in this would be an openness to novelty since, as just said, harmony is not predetermined.

According to Chenyang Li, interpersonal harmony also requires an accommodating attitude, a tentative acceptance of "elements that are not yet conducive to harmony" (2014, p. 126). Accommodation is crucial because even though a person's uniqueness is accepted, this does not mean the person does not need

---

69 The passage reads as follows: "Harmony is like a stew. Water, fire, jerky, mincemeat, salt, and plum vinegar are used to cook fish and meat. These are cooked over firewood. The master chef harmonizes them, evening them out with seasonings, compensating for what is lacking, and diminishing what is too strong. The noble man eats it and calms his heart" ("Zhaogong" 20.8a in *Zuozhuan*, as translated in Schaberg et al., 2016, p. 1587).和如羹焉, 水、火、醯、醢、鹽、梅, 以烹魚肉, 燀之以薪, 宰夫和之, 齊之以味, 濟其不及, 以洩其過. 君子食之,以平 其心

70 In Li's account, he translates *geng* 羹 as "soup."

to change to harmonize with others. Harmony is about balancing elements, after all, and so all parties concerned must undergo some change to work harmoniously. However, change does not happen immediately, so other group members should be patient enough to accept those who still need to integrate themselves. The elements in *League*'s ranked games can hinder the players' development of the attitudes of openness and accommodation. Recall that the ranking system affords an intense motivation on the part of players to win matches and rise in rank. The snowball effect also affords a scenario where players are excessively wary of their team making even a single mistake since that might be enough to lose the game. Combined, these two affordances create a compelling demand to stick rigidly to conventional playstyles to ensure optimal game performance while making it difficult for players to tolerate different or novel playstyles for fear of losing. In fact, players are known to act toxic towards teammates who use unconventional playstyles (DongHuaP 2021, 7:50-55), and some even quit the game once they realize that a teammate is playing a champion outside the established norm (2021, 3:19-24).

This situation is adverse to the cultivation of openness because it predisposes players to be sceptic of differences and novelty. But these circumstances are also unconducive for cultivating the attitude of accommodation since players cannot tolerate people who are new to ranked matches and require some time to integrate themselves. An example can be seen in the quote from section 3.2, where a beginner in ranked games was immediately ridiculed and blamed for causing the team's loss after only being killed once.

This section has shown that the elements in ranked games and the toxic behavior they afford create an unconducive environment for cultivating fundamental Confucian values. However, Confucians were no strangers to living in such environments. It is believed that the ancient Confucian masters lived during less-than-ideal times when the socio-political circumstances were hostile to cultivating the virtues they espoused. Due to this, they needed to think of means to develop ethical attitudes. In the next section, we shall discuss three of these means and explore how they can inspire concrete proposals to remedy *League*'s toxicity.

3.4. Confucian-inspired Measures to Lessen Toxicity

The ancient Confucian masters thought of several means by which people could achieve promoral or virtuous dispositions. This section (3.4.) discusses three of them

and explains how these ideas can serve as inspirational resources for developing measures to remedy toxicity in *League*.

3.4.1. Redefining Characters' Defeats

The idea of utilizing or "channeling" one's non-moral inclinations to achieve moral ends is derived from Mencius. It can be viewed as his practical method for encouraging the non-virtuous to begin the process of moral cultivation, even though they have desires that (could) lead to vice. In Mencius' dialogues with King Xuan, King Xuan confesses to Mencius that several inclinations prevent him from becoming virtuous. Examples are the fondness for inappropriate music, courage, wealth, and sex.[71] In each case, Mencius does not tell King Xuan to stop entertaining such inclinations. Instead, he urges the monarch to channel his tendencies toward realizing moral goals. For example, Mencius tells the king to harness his fondness for courage by fighting injustice and bringing peace to the world, not by avenging himself against petty insults. Similarly, Mencius instructs him to use his passion for wealth to enrich not only himself but also his people.

The idea of channeling can be applied to *League* to address the game's valorization of killing. As discussed in section 3.3.1, *League*'s valorization of killing can dispose players to trivialize human life and glorify killing. This can be problematic even from a non-Confucian perspective since it might promote aggressive and antisocial behavior. However, killing champions in *League* is not generally done for its own sake. Instead, players kill champions to display their skillfulness and attain a high rank. Thus, we propose to remove killing from ranked games and replace it by portraying a character's defeat not as a "death," but simply a "retreat." More specifically, a defeated character can be represented as merely becoming weak and then "teleporting" out of the map. Furthermore, instead of announcing champion defeats as "kills," they can be stated as "wins" (on the part of the player who defeated the champion).

Replacing the element of killing in ranked games with the one of defeating will not repress players' inclination to display their skillfulness but channel it to a more morally acceptable end. Indeed, introducing the element of defeating does not deny players the opportunity to show their skills but ensures that they pursue this inclination towards the realization of a different goal. Admittedly, the goal of "not killing" is not

---

[71] *Mencius* 1B.1:18, 1B.3:48, 1B.5:4, 1B.5:5, as translated in (Mencius, 2008, pp. 16-17, 19, 22-23).

necessarily moral. Thus, strictly speaking, we propose channeling players' non-moral inclinations from an immoral end (killing) to a non-moral one.

One could object that removing killing will not eliminate aggressive or antisocial behavior because the rules still require champions to hurt each other, and thus commit violence. Although this might be true, we believe that our recommendation is a promising first step towards that goal. Further steps can be carried out by our two other recommendations that aim to instill moral and prosocial dispositions in players.

Another way of channeling players' inclination away from the valorization of killing concerns the kind of actions recorded during matches. The game records the number of times a player has killed an opponent and assisted an ally in getting kills. The number is visible for players both during and after a match. However, this tracking of life-taking actions glorifies killing and can suggest that lives are not noteworthy or valuable. To remedy this, we propose the game records lifesaving acts, such as when players protect their teammates from being defeated by opponents. By doing this, the game can instill an appreciation for life rather than killing in players.

3.4.2. Re-Watching Feature

From a Confucian perspective, a lack of sympathetic concern (*shu* 恕) for players can contribute to toxic interaction between teammates. Furthermore, as we argued in section 3.3.2, one of the most severe obstructions to cultivating the attitude of sympathetic concern in *League* is the attention-demanding nature of games. At a minimum, sympathetic concern requires attention and time to imagine other individuals' thoughts and feelings. Still, *League*'s players who are seriously playing a match are often unable to make space for it.

We propose to remedy this problem by giving players sufficient time and "attentional space" to put themselves in the place of others. Specifically, we suggest introducing a new feature in the game that allocates time to players who have been reported for toxic behavior to reflect on their behavior towards other players.[72] At the end of a game, the feature will invite the player reported for toxic behavior to rewatch the highlights of their interactions with her teammates during matches. To help the player imagine herself in her fellow teammates' place, the recording will show the players' interaction from the standpoint of the players who have filed the complaint. This will allow the reported player to visually and aurally represent what her

[72] The report system allows players to report other players who express toxic behaviour. See: Riot Games (2022b), Reporting FAQ section.

teammates have experienced while interacting with her. If the invitation mechanism proves unsuccessful, the feature can be made compulsory so that reported players must watch the highlights of their previous game to continue playing matches. Furthermore, after watching the matches' highlights, the player may be asked questions like "What do you think the other player felt because of your behavior?" and "If you were in his place, how would you want to be treated?"

3.4.3. A New Interpersonal Practice to Foster Players' Reciprocal Respect and Appreciation

For the ancient Confucian masters, *li* 禮, or "rituals," were a set of individual and social practices to regulate human relationships at different societal levels. At a minimum, rituals contribute to creating cordial interpersonal relationships. They serve as codes of conduct and social customs that prescribe the proper behavior for different situations (Lewis, 2020, p. 43). Additionally, rituals are instrumental to personal moral cultivation because they specify the appropriate emotions and moral dispositions to inform such actions (2020, p. 78). For instance, Xunzi says that the funeral ritual resembles one's moving to another house to increase sorrow for the deceased's departure.[73] This shows that the mourning ritual is not just about the performance of specific actions during the funeral but also the development of a particular emotional disposition in the person who participates in the ritual. Notably, the practice of ritual was not expected to be mindless repetition. For instance, Xunzi emphasizes the importance of reflecting on "what is central to ritual"[74] to achieve moral perfection.[75] Similarly, Confucians believed that if one practiced rituals regularly and reflectively, one could come to appreciate their underlying ethical principles and develop an intrinsic motivation to enact them (Lai, 2016, p. 29; Cline, 2016, pp. 18-19; Ivanhoe, 2014, p. 49).

This idea of Confucian ritual can inspire new solutions for combating toxicity in *League*. As implied in section 3.3.2, toxic behaviors in ranked games are primarily rooted in players' intense and individualistic motivation to win a high rank. Players are motivated to attain a high rank so strongly that they disregard respecting their fellow players to pursue it. A possible remedy is introducing a countervailing

---

73 *Xunzi* 19: 414-415, as translated in: (Xunzi, 2014, p. 212).

74 The phrase is taken from the sentence: **禮之中焉能思索, 謂之能慮**. "To be able to reflect and ponder what is central to ritual is called being able to deliberate" (Xunzi, 2014, p. 206).

75 *Xunzi* 19: 165-170, as translated in: (Xunzi, 2014, p. 206).

interpersonal practice inviting players to foster more respectful and friendly relationships with fellow players. This opposing disposition can be promoted through an in-game feature that helps diminish players' obsession with high rank and increase their appreciation for other players. Our proposed feature is inspired by Confucius' description of the ritual that the *junzi* 君子 ("gentlemen" or virtuous persons) performed during archery contests. The ritual requires junzi to show respect and fellowship to each other before and after the contest. "Before mounting the stairs to the archery hall, gentlemen [*junzi*] bow and defer to one another, and after descending from the hall they mutually offer up toasts."[76]

Inspired by this passage of the *Analects*, we propose to develop a new feature in *League* that requires each player to give at least one positive message to one of her teammates and one to an opposing team player before and after each match. The exchanged message before the match would be a word of encouragement or friendliness, while the message afterwards could be praise or constructive criticism. To help players express openness and accommodation, the feature can even suggest players commend other players for their unique playstyles or give constructive criticism to those who seem to be beginners.

Similar to the idea of Confucian rituals described above, the proposed feature would not only indicate to players a course of action but encourage them to display the appropriate emotions while doing the action. In the context of *League*, this would be a respectful and friendly spirit proper to a sport or game. One way the feature can induce this spirit is by having the procedure's instructions narrated by a warm and welcoming voice. This could influence the mood of the players when they relay their messages. Furthermore, the visual design of the feature can also evoke a sense of camaraderie and fellowship. Finally, to invite players to reflect on the purpose of this practice, the feature will inform players that the procedure aims to help them be respectful and kind to each other while playing *League*.

3.5. Conclusion

Video games have become a relevant aspect of contemporary societies in the past four decades, acquiring "cultural and social relevance" (Muriel and Crawford 2018, p. 22). This raises questions about the cause of popular video games' toxicity and the possibilities to redress such a phenomenon. This paper advances the understanding of

[76] 揖讓而升, 下而飲. *Analects* 3.7, as translated in: (Confucius, 2003, p. 19).

these two issues through a "Confucianinspired approach" (Dennis & Ziliotti, 2022). Focusing on the multiplayer video game *League of Legends* as a case study, we have argued that *League*'s affordances can hinder the cultivation of fundamental Confucian moral dispositions. However, Confucianism is a rich source of inspiration for conceptualizing possible remedies for *League*'s toxicity. First, the idea of using a person's nonmoral inclination to lead them to do moral (or non-immoral) actions can help remedy for *League*'s valorization of killing, which is problematic from both a Confucian and non-Confucian perspective. Second, the Confucian disposition of *shu* 恕 (sympathetic concern) suggests the possibility of inviting hostile players to rewatch the highlights of the match in which they exhibited negative behavior. Finally, a code of conduct constitutive of the notion of Confucian *li* 禮 (ritual) can inspire new scripted procedures to foster players' mutual respect.

This paper provides a significant opportunity to advance Confucian-inspired ethical studies on technologies by bringing attention to the moral dimension of video games, a relatively underexamined topic. Furthermore, it suggests a new course of action for Confucian-inspired ethical studies of video games. While this paper has focused on *League*, online games' toxicity depends on their distinctive affordances. Thus, a key priority for future studies is to explore whether and how Confucian ethics can contribute to mitigating those forms of toxic online gaming behavior. Finally, certain Confucian virtues, such as mutual respect and sympathy, appear quite needed in our world today. Thus, another direction for future Confucian studies of video games is to explore how insights from Confucian ethics can help game designers who are working to unlock the potential of video gaming for engendering these important dispositions.[77]

[77] Psychologists have begun to pay attention to the effects of game playing on prosocial character and with interesting results. Jamil Zaki discusses some of these works, particularly those involving gamified applications. See: (2019, p. 144-167).

Interlude 3

The previous chapter examined how Confucian philosophy can provide design recommendations for online multiplayer games, such as *League of Legends*, that have the potential to reduce toxic behavior between players and, thereby, contribute to the pursuit of the digital good life in the interpersonal sphere. The next chapter will address online political polarization in social media. Social media often fosters polarization, aligning individuals with specific political groups and generating distrust among citizens. This animosity can undermine the social bonds and cooperation essential for effective political governance and community well-being (cf. Ziliotti 2022). Thus, the issue of political polarization in social media can be considered as belonging to the sphere of the political. The following chapter will examine how Confucian philosophy can serve as a potential source of insights for reducing online polarization. Since emotion plays a significant role in online polarization, the chapter will examine ideas and methods in Confucian philosophy that utilize emotions to promote attitudes that are inimical to polarization. Confucian philosophy is a promising resource for this, as it values certain emotion-laden attitudes, such as care and sympathy, that are, in turn, inimical to polarization. By applying Confucian philosophy in this manner, the chapter shall demonstrate how it can contribute to the pursuit of the digital good life in the political sphere.

Chapter 4. Mitigating Online Polarisation from the Perspective of Confucian Cognitive Affectivity[78]

4.1. Introduction

Online polarization (OP for short) is a prominent problem. A common idea is that this is brought about by an ever-growing divergence of opinion facilitated by filter bubbles and echo chambers and ultimately rooted in the exercise of users' rational deliberative capacities. An alternative to this idea is the theory of "affective polarization," which considers OP as a function of the emotionally-laden ties users have with their in-group, which conversely translates into animosity for the partisan outgroup and their policy preferences. Despite rightly identifying the role of emotions in OP, the theory of affective polarization is nevertheless in danger of framing emotions as mere fuel for polarization and, thus, needing to be eliminated or dismissed to mitigate polarization. In this paper, we will challenge this assumption by discussing how cognitive theories of emotions can provide a different approach to addressing the problem of OP, one that considers emotions as having an important role in lessening polarization and promoting mutual understanding. To support our argument, we will look to insights in Confucian philosophy that emphasize the role of cognitive emotions in moral development. We will also draw inspiration from Confucian methods of orienting emotions towards ethical ends to recommend strategies for mitigating OP in ways that precisely utilize the emotions instead of dismissing them.

A popular way to explain OP has been in terms of filter bubbles and echo chambers. The filter bubble theory states that the personalization algorithms of online platforms, such as social media, expose users only to the content they prefer (Spohr, 2017). This prevents users from having a more holistic and accurate perspective of issues. This, in turn, contributes to polarization since users become more entrenched in their opinions, which concomitantly makes them more dismissive of people who have opposing views. Moreover, filter bubbles make it easier for users to be exposed only to like-minded people. This can lead to the formation of echo chambers, that is,

[78] This chapter is based on a paper co-authored with Sabine Roeser, with the author of this dissertation as the first author. The co-author has given her permission to use this paper for the dissertation.

communities that discredit the epistemic merits of non-members while at the same time valorizing their own (Nguyen, 2018). Echo chambers induce their members to trust only the information their fellow members provide, which can result in biased perspectives and polarization between the community and the groups it treats as adversaries.

One strategy to mitigate OP is designing digital design features to diversify the online content to which users are exposed. This has been called by Nelimarkka et al. (2018, 2019) as the "Common Design Agenda" (CDA for short) and is popular among HCI researchers. The CDA attempts to provide users with a more balanced and less partisan perspective of issues, thereby supposedly mitigating the biased view filter bubbles and echo chambers effectuate. Some studies, however (Liao & Fu, 2013; Bail et al., 2018; Weeks et al., 2017), reveal that exposing users to information from partisan outgroups can actually increase polarization.

We believe the CDA's ineffectiveness is due to an implicit rationalistic bias that informs it. By considering diversifying information as the means for mitigating polarization, the CDA appears to assume that users interact with this diverse information in a disengaged and impartial manner. Users are tacitly viewed as detached rational actors that, by default, would neutrally assess the diverse content presented to them without any emotional influence. In other words, this rationalistic bias downplays or ignores the affective elements involved in users' exposure to diverse content.

Certain frameworks, however, have highlighted how emotions play a role in users' experience of social media (e.g., Marin & Roeser, 2020; Marin, 2024). For instance, Steinert et al. (2022) have argued that social media platforms are environments with design and social elements that influence users' emotions and, consequently, their beliefs, attitudes, and actions. Moreover, as we will show in section 4.3, the theory of affective polarization highlights the critical role emotions play in generating polarization. Thus, to address problems like OP, emotions should be given greater attention and appreciation.

This paper aims to build on this insight to counterbalance the rationalistic bias in approaches and strategies for mitigating OP. This paper will suggest that paying attention to emotions can play a crucial role in helping steer users away from polarization and toward more harmonious interaction with others. To support this, we will draw on cognitive theories of emotions that emphasize the importance of emotions for moral interaction and mutual understanding. We will also draw on the resources of Confucian philosophy to further concretize how the notion of cognitive

emotions can be used to reduce polarization and promote harmonious interaction. Confucian philosophy pays significant attention to emotions because it considers them an ineluctable part of human life and ethical flourishing (Nelson, 2018). Moreover, certain strands of thought in the Confucian tradition consider emotion as inextricably bound up with cognition. Confucian philosophy can, therefore, be a promising source of ideas for mitigating polarization that take into account and utilize cognitive emotions.

We proceed as follows. In section 4.2., we discuss the rationalistic bias in addressing OP as it exists in different approaches to and conceptions of OP. We then indicate the shortcomings of these approaches. After this, we describe, in section 4.3, an alternative conception of polarization, namely, affective polarization, and show how this conception fits better with empirical studies. Although the idea of affective polarization indicates the importance of emotions in OP, it could also give the impression that emotions are wholly negative elements and, therefore, should be dismissed or eliminated when mitigating OP. We counter this notion in section 4.4., where we discuss cognitive theories of emotions and emotional deliberation to show how emotions have ethical importance and why they should be considered positively when addressing OP. After this, in section 4.5., we look to Confucian philosophy to further illustrate the ethical potential of cognitive emotions. We show how Confucian philosophy also has a cognitive understanding of emotions and how these are crucial for ethical flourishing. Moreover, we cite specific Confucian ideas for utilizing cognitive emotions for ethical ends. In section 4.6., we draw inspiration from these aforesaid ideas to articulate concrete methods of mitigating OP that use emotions instead of dismissing them. We then conclude the paper with a summary in section 4.7.

4.2. Rationalistic Bias of Standard Approaches to OP

Several approaches to framing and mitigating OP are grounded in a rationalist approach. We will argue in this section that this can be considered a "rationalistic bias" in that it ignores emotional or affective elements that come into play in the aforesaid phenomenon. Consequently, they also overlook the constructive contribution emotions could play.

We can begin by discussing the approach already mentioned above, namely, the CDA. Again, the CDA intends to design digital interventions that 'aim to balance and diversify information and news' (Nelimarkka et al., 2019, p. 2) that users are exposed

to. An example is Chhabra and Resnick's 'Cubethat,' a browser extension that shows users links to stories related to the one they are currently reading (2012). Another, more recent, example is Garimella et al.'s attempt to bridge echo chambers through an algorithm that would introduce content from a user in one echo chamber to a user from another echo chamber (2017). The CDA assumes that exposing users to diverse content, particularly those from the opposite groups, can lessen OP. This is because such exposure could free them from the myopic views afforded by filter bubbles or echo chambers and the antagonistic attitudes toward the outgroup that these views engender. The CDA can be considered as having a rationalistic bias in that it seems to assume that users will, as a matter of course, impartially consider alternative content and adjust their attitudes accordingly. It does not consider the possibility that users might react differently to this content due to their affective dispositions. Furthermore, they overlook the possibility that emotions could play a constructive role in overcoming the OP. This latter point will be the main focus of the alternative approach in the following sections.

To further elucidate the rationalistic bias of the CDA, we can look at the conception of polarization that arguably informs it. Törnberg et al. coin this as 'opinion-driven polarization' (2021). The theory of opinion-driven polarization assumes that polarization is a function of the ever-widening divergence of policy preferences or opinions that users have and that this divergence is ultimately rooted in the exercise of their rational deliberative capacities (cf. Kim & Kim, 2019; Buder et al., 2021). The theory explains the process of OP as follows: Users first come to prefer certain opinions or policies through rational deliberation. The personalizing algorithms of new media technologies, such as social media, then consider these preferences. This leads to the creation of filter bubbles and echo chambers, which ensure that users are exposed only to preferred content and like-minded people. Within these bubbles and chambers, users lack exposure to information that might counterbalance or challenge their beliefs. Consequently, users become more extreme in their convictions. This eventually leads to polarization because groups are formed based on drastically and increasingly differing opinions, and the growing gap between these groups fosters animosity and a breakdown of civil discourse. The remedy to this is to expose users to information beyond their own bubble or chamber — which is precisely what interventions based on the CDA attempt to do.

Related to the CDA and the idea of opinion-driven polarization is another approach to addressing OP that arguably illustrates a rationalistic bias. Political philosopher Elizabeth Edenberg calls this the "epistemic response" to OP (2021). The

primary solution that the epistemic response provides to OP is the enhancement of users' epistemic capacities and attitudes. This can be seen, for instance, in the attempt of both social media companies and educational institutions to promote initiatives for teaching people how to evaluate online content responsibly (Edenberg, 2021, pp. 262-263). Scholars have also argued the need to cultivate epistemically virtuous attitudes, such as intellectual humility and avoiding epistemic laziness, so that users can transcend their bubbles or chambers (Lynch, 2021; Nguyen, 2018). For Edenberg, the difficulty with the epistemic response is that it ignores non-epistemic factors that contribute to OP, such as social media algorithms and social ties (2021, pp. 267-268). Although Edenberg does not explicitly state it, the epistemic response can also be viewed as ignoring the emotional dimension of OP. For instance, suggesting that the way out of filter bubbles can be achieved by overcoming one's epistemic laziness assumes that such laziness is the main cause of users' avoidance of alternative information. It does not consider that users avoid alternative information, not out of laziness, but out of *unwillingness* due to their affective ties to their partisan group and concomitant hostility towards the partisan outgroup.

All these approaches, then, focus on information and rational deliberation. However, they largely ignore the affective or emotional elements that play a role in OP. That there is something amiss with the rationalistic bias of these approaches can be gleaned from empirical studies whose results cast doubt on some of the important assumptions and claims of these approaches that stem from their rationalistic orientation. First, studies have cast doubt on the existence of filter bubbles and echo chambers, or, in other words, the idea that users can be thoroughly insulated from alternative information (Barberá et al., 2015; Bruns, 2019; Zuiderveen Borgesius et al., 2016; Dubois & Blank, 2018; Scharkow, 2020; Nguyen & Vu, 2019; Nechushtai & Lewis, 2019). Second, studies have shown that a radical divide in political opinion does not occur among the citizenry (Gentzkow, 2016; McCarty, 2019; Munroe, 2023; Bail, 2021). Instead, studies have shown that "the political positions of most voters remain moderate" (Törnberg et al., 2021, p. 3). This challenges the claim of opinion-driven polarization that such a radical divide exists (due to filter bubbles and echo chambers). Third, studies have shown that users' exposure to diverse views does not necessarily reduce polarization (Lee et al., 2014; Liao & Fu, 2013). Indeed, some studies suggest that exposing people to opposing views online *increases* polarization either by engendering feelings of hostility among users in relation to the outgroup (Bail et al., 2018) or by prompting users to purposely seek out preferred or pro-attitudinal content (Weeks et al., 2017). This directly challenges the reasoning behind the CDA,

which is that diversifying content can reduce polarization. This also raises the question of whether developing users' epistemic capacities to get them to leave their bubbles, as the epistemic response recommends, is the main solution to OP. Finally, some studies have shown that OP is moderated by the level of civility and the sort of sentiment in which online information is packaged rather than the information itself (Kim & Kim, 2019; Buder et al., 2021). This challenges the rationalistic assumption that users consider online information in a detached or neutral manner.

4.3. Affective Polarization as an Alternative Explanation to OP

Törnberg et al. refer to a theory that can serve as a foil to the rationalistic biases of the abovementioned approaches in that it considers emotions as having a major role in OP (2021). This is the theory of affective polarization. Unlike the theory of opinion-driven polarization, affective polarization claims that partisans generally express negative feelings toward the out-group, regardless of ideological differences (cf. Iyengar et al., 2019; Talisse, 2019; Arora et al., 2022). This means that members of a group exhibit negative dispositions towards members of opposing groups even though the beliefs of both groups do not significantly differ. Affective polarization also introduces a different way of explaining OP compared to opinion-driven polarization. Instead of the root of OP being diverging opinions, it is the strong affective ties users have to their partisan group. Because users consider their membership in their group as a salient part of their identity, they tend to form a robust affective attachment to it, as well as its members, beliefs, and values (cf. Munroe, 2023; Marchal, 2022; Nordbrandt, 2021). Conversely, users are disposed to express intense emotional dislike for opponent groups and what they stand for. This, in turn, leads users to adopt and defend their group's opinions or policies (Törnberg et al., 2021). In other words, the divergence of opinion is the function of polarization rather than its basis.

In addition to these observations, we would like to add that these groups may be more value-based identities and groups than traditional local 'clans,' i.e., I may feel more connected to someone with similar values and viewpoints from the other side of the globe than, say, with my neighbor who holds conflicting values to mine. We can, in recent years, also see that nationalist politicians from around the globe find each other in a paradoxically international nationalist movement. As Kahan (2012) has shown, people find experts who share the same values as they do more credible. Thus, values and concomitant emotions can be a causal factor in the kind of information that people find credible. This goes against rationalist assumptions, which state that

people first take in neutral, objective, factual information and, based on that, form their value judgments, which are directly and unambiguously determined by the facts. Still, polarization may not be irrational and arbitrary, but it can also be grounded in deeply held values and related emotions. Hence, we would like to add to Törnberg et al.'s work that affective polarization not only has an emotional dimension related to local identities but also a value dimension related to value identities.

This view of affective polarization, grounded in values, also sits well with the empirical findings above because it does not assume that lack of exposure to out-group content is necessary for polarization. It also does not entail a dramatic divide in opinion among the general citizenry, as with opinion-driven polarization. However, it cautions that OP may spread from the extreme fringes to the more nuanced population, as people may feel a strong need to protect their core values based on strong appeals to these values from the more extreme positions that may be more vocal and visible online, due to the enforcement mechanisms on many social media platforms that tend to push these kinds of posts, as they provide more 'clickbait'. Furthermore, affective polarization can explain why exposing users to alternative content does not always work and why the CDA might not be entirely effective. Instead of dispassionately and impartially considering the diverse content presented to them, users might instead be disposed to act hostile towards these due to feelings of defensiveness toward their group and core values, and antagonism for the out-group and their values (from which these contents originate). For similar reasons, the theory of affective polarization also shows why the epistemic response might be inadequate. The epistemic response does not consider users' affective ties as a contributing factor to OP. It assumes that OP is merely the outcome of the inability of users to exercise epistemic virtues and attain epistemic goods.

The theory of affective polarization suggests, then, that the abovementioned approaches that admit to a rationalistic bias (i.e., CDA, opinion-based polarization, epistemic response) fail to diagnose OP in a wholly accurate way because they construe OP mainly as a (dys)function of people's rational and epistemic capacities. They fail to highlight that OP can involve in-group and out-group dynamics (also related to communities of shared values that can be globally dispersed), which are facilitated by online platforms and the affective elements that constitute it. We do not mean to deny that the rationalist assumptions underlying the CDA and other approaches may be adequate in some contexts. However, such approaches overlook the crucial role that emotions and values can play in OP. As for the theory of affective polarization itself, although it rightly identifies emotions as playing a significant role

in OP, there is a danger that it might frame emotions simply as a negative force that drives polarization and, thus, something that must be overcome or eliminated. This, in turn, precludes the possibility of considering emotions as a means itself of mitigating OP. In other words, interventions aimed at mitigating OP could potentially be improved if they take advantage of the affective and value elements of OP in addition to the ones related to factual information that are supposedly a matter of detached, rational engagement. However, this potential cannot be actualized if such elements are considered things to be eliminated or dismissed.

In what follows, we will build on and expand these insights. We intend to use ideas from Confucian philosophy as inspiration for the abovementioned interventions. This is because Confucian philosophy has a considerable appreciation for emotions and their normative significance. It also has ideas for utilizing emotions to help people achieve ethical ends. Moreover, Confucian ideas also provide the foundations for a strong ethical appeal to designers of social media platforms to reconsider their profit models by emphasizing their moral emotions, virtues, and leadership. Before this, however, we need to identify what exactly it is about emotions that allow them to play the positive role of mitigating OP and why it is better to actualize such a role for ethical and practical reasons. To do this, we shall discuss the notions of cognitive emotions and 'emotional deliberation' in the next section. We will also highlight the importance of values in online deliberation, something that is not prominent in the approaches we discussed in this section, which primarily focus on the exchange of and access to descriptive, factual information. These discussions in section 4.4. will, in turn, provide the conceptual backdrop for the Confucian ideas to be tackled in section 4.5.

4.4. Emotional Deliberation as an Alternative View for Mitigating OP

In section 4.2, we saw that leading approaches to address OP are grounded in rationalist ideas of providing access to factual information from diverse sources. This is particularly visible in the CDA approach but is implicit in the theory of opinion-driven polarization and the epistemic response. We analysed this as a 'rationalist bias.' The rationale behind this 'bias' is a common idea about the supposedly opposing roles between reason and emotion. Emotions are usually seen as the culprit in polarized debates. They are seen as irrational factors that prevent people from rationally processing objective, value-neutral information that would settle the debate. This idea is also reflected in the highly influential 'Dual Process Theory', according to which people process information through two fundamentally different systems. System 1 is

supposed to be emotional, intuitive, irrational, and spontaneous. It operates quickly but at the cost of being highly unreliable. System 2 is supposed to be rational, analytical, and sophisticated, but it is time-consuming and laborious. System 1 is supposed to be evolutionary prior to system 2 (cf. e.g., Kahneman, 2011). This framework has been very influential within social psychology and decision theory, and it also aligns with the dichotomy between reason and emotion prevalent among many philosophers. In public debates, this dichotomy is also frequently used to dismantle one's opponent. By calling someone emotional, they are framed as irrational, unscientific, and simply wrong. These ideas are also implicitly at play in analyses of polarization in social media, as we saw in the previous section. For instance, the common assumption that users will dispassionately engage with diverse or alternative information arguably valorizes detached rationality or system 2 as that which is useful for overcoming conflictual issues like OP.

In what follows, we will challenge this analysis by debunking the reason-emotion dichotomy, introducing the ideas of affectual intuitionism, emotional deliberation, and the importance of emotional attitudes towards moral exemplars. In section 4.5, we will discuss how insights from Confucian philosophy can further strengthen this cognitive understanding of emotions by providing a stronger ethical dimension for said emotions. We will also present ideas from Confucian philosophy that utilize the cognitive understanding of emotions for ethical ends and which can serve as inspirations for methods of mitigating OP that precisely make use of emotions instead of ignoring them.

In contrast with the dichotomous view of reason and emotion sketched above, emotion researchers in psychology and philosophy have, for the last decades, provided alternative approaches that challenge this dichotomy between rationality and emotion. So-called cognitive theories of emotions emphasize the cognitive aspects of emotions, next to affective and motivational aspects (e.g., Scherer, 1984; Lazarus, 1991 in psychology, and Solomon, 1993; Nussbaum, 2001; and Furtak, 2019 in philosophy). Neuropsychological research has shown that people with specific forms of brain damage to their amygdala lose their capacity to feel emotions and thereby also their capacity to make practical and moral judgments in concrete circumstances, while their IQ (or analytical rationality) is unaffected (Damasio, 1994). Interestingly, these ideas can already be found in ancient Western philosophy, such as in the works of Aristotle and classical Chinese philosophy. In classical Chinese philosophy, the notion of 'heart-mind' is crucial, which defies the reason-emotion dichotomy and is similar to the cognitive theories of emotions mentioned above. We will discuss this last

point in more detail in section 4.5. In any case, these insights challenge the reason-emotion dichotomy prevalent in research and public debates. In the remainder of this section, we will present an alternative theoretical framework that shows how emotions can play an important role in overcoming polarization. This option is largely overlooked due to the prevalent dichotomous framework.

In what follows, we will draw on the theory of 'Affectual Intuitionism'. 'Affectual Intuitionism' is a philosophical theory that combines ethical intuitionism with a cognitive theory of emotions. It argues that emotions are value judgments through which people perceive or 'experience' moral value. Affectual intuitionism builds on the theory of ethical intuitionism by acknowledging the importance of particular, context-sensitive ethical judgments (Roeser, 2011). However, ethical intuitionists are typically rationalists; they endorse the reason-emotion dichotomy sketched above. Some of them allow emotions a role to provide for motivation (e.g., Reid 1788/1969), but they do not allow emotions an epistemological role (cf. Roeser 2011, part II, for a critical discussion). Affectual intuitionism rejects the reason-emotion dichotomy and instead builds on a cognitive theory of emotions. It understands moral emotions as felt value-judgments (Roeser, 2011). Emotions thus understood can then function in the way that intuitions do for ethical intuitionists: ethical intuitions understood as cognitive moral emotions are not deductive, inferential, or strictly argumentative. Rather, paradigmatically moral emotions let us assess the moral value of a situation in a direct, experiential way, analogous to perception. Furthermore, they can ground our further moral reasoning, in a similar way as intuitionists have argued concerning intuitions. Emotions are better suited for such context-sensitive judgments than reason since emotions are especially sensitive and attuned to morally relevant aspects of concrete circumstances (Roeser, 2011). People can be led to perceive and appreciate values if they experience emotions that 'disclose' such values. This can open the way to a more shared understanding of different viewpoints. Moreover, affectual intuitionism says that people are motivated to enact the value precisely because of the emotion they feel about it.

This is crucial as what is often overlooked is that polarized debates online, which also spill over into the 'offline' world and vice versa, are not just about descriptive facts but also about values. This is where the epistemological role of emotions comes in: not by assessing the reliability of scientific data – this needs to be done with scientific methods. Rather, by highlighting evaluative aspects that are intrinsic to many, if not all, debates about controversial issues. The assessment and implications of scientific research for public policies and decision-making are not value-neutral. Rather,

science provides necessary but insufficient information for decision-making and policy options. Based on scientific information, options for action arise that are scientifically underdetermined. Issues such as justice, autonomy, and fairness are typically involved in scientifically informed policy options. These issues cannot be solved by scientific means, as they are not descriptive but evaluative, requiring an ethical point of view. This is where emotions come in. On the theory of emotions sketched above, emotions are crucial for such a moral point of view (cf. Roeser, 2018). This does not reduce ethics to mere subjective attitudes, as is the underlying assumption in the rationalism-sentimentalism debate in metaethics. Rather, grounded in a cognitive theory of emotions, affectual intuitionism argues that emotions are perceptions of values. If values depend on emotions, that would be circular and would not help explain why some situations give rise to one kind of emotion and other situations to others. While sentimentalists such as Humeans see emotions as subjective projections on a morally neutral world, affectual intuitionism argues that emotions make us sensitive to evaluative aspects of the world.

Emotions can then help us to be sensitive to the ethical dimensions of complex decision-making problems. For example, indignation can highlight aspects of injustice or violation of autonomy, and enthusiasm can highlight the positive impacts of a decision option. Care and sympathy can help to understand the moral implications of effects on others, as well as help to understand their viewpoints, by putting ourselves in somebody else's shoes. In other words, emotions can contribute to ethical deliberation, what has been called 'emotional deliberation' (Roeser & Pesch, 2016; Roeser, 2018). Strikingly, we find this idea, or something highly similar to it, in the work of Mencius, as we will discuss in section 4.5.

This proposal to give more importance to emotions in ethical deliberation, online or offline, may come as a surprise, as typically, emotions are blamed for polarized debates. However, this overlooks the rich spectrum of emotions. Emotions are not only antagonistic, as in the case of aggression, anger, and hatred. Rather, pro-social emotions such as care, empathy, sympathy, and compassion are crucial for reflecting on ethical issues and trying to understand another person's point of view. In other words, such emotions can help to focus on what unites us, instead of what divides us. These emotions can thereby help provide more nuance, which can take away the fuel for polarization. This is not to say that scientific information is irrelevant. On the contrary, as said, it is necessary for decision-making on many complex issues. However, as was also mentioned, this information is probably not sufficient, as there are typically ethical issues involved on which emotions can shed important light.

Moreover, evading the evaluative dimension can fuel emotions, as people can feel that their concerns are not being heard. People may then find recourse in misinformation that they think supports their values, as that misinformation may present itself in the form of science and may suggest to be more legitimate in a science-dominated discourse. By opening complex debates explicitly to not only scientific but also ethical issues and allowing emotions in the arena to highlight the latter, the pull towards bad science may become less strong. This, then, can lessen the risk of further polarization, thereby making the strategy of entertaining emotions more effective and practical in precisely mitigating OP.

Even if the importance of deliberation on values is acknowledged in ethics research on OP, the emphasis is usually on rational arguments. However, that can miss important nuances and understanding. More seriously, this exclusive focus can marginalize those who are "not trained to frame their […] deliberations in such a way as to comply with disengaged rationality" (Roeser & Pesch, 2016, p. 282). This may include people of lower income brackets who might not have been privileged with an education that precisely teaches them such framing. As such, allowing emotions in deliberation can also be considered a form of justice since it allows the voices of those who could potentially be excluded by a solely rationalistic method to be heard. Furthermore, rationality can also be fallible: it can emphasize self-interested decisions, which can be suboptimal and morally problematic, for example, as in so-called free-riding behavior and 'prisoners' dilemmas' (cf. rational choice theory). Such biases of rationality can be corrected by moral, other-regarding emotions. Hence, all kinds of knowledge, be it grounded in rationality, emotions, or a combination of the two, require critical reflection and deliberation, and these forms of knowledge can themselves play a role in such critical deliberation. The plea for emotions does not mean that emotions are infallible. To the contrary, they need to be reflected upon, critically examined, and educated. However, emotions can play an important role in critical reflection through 'emotional deliberation.'

Purely cognitive moral judgments without any emotions involved have a hard time fully capturing moral knowledge. For example, if we read in a newspaper that a tragedy has happened somewhere in the world, we will typically judge that this is terrible without strong emotions. However, if we start to see pictures or videos from that event, for example, in an online post, or even more so if we are a direct witness, let alone a victim, we will typically feel intense emotions.[79] These emotions can also

[79] This connects with Mencius' famous example of someone seeing a child falling in a well, see section 5.1.

lead to a more thorough understanding of the complex ethical aspects, and they provide us with genuinely lived-through moral judgments. This can be the strength of emotional involvement on social media. However, the downside can be that we only see one side of the story and miss further details and context. Still, enticing people to be prepared to also take in other information and viewpoints that may make matters more complex is not just a matter of appealing to rational arguments. Rather, it requires an emotional openness, empathy, and respect for others to first engage with a fuller picture before jumping to conclusions. Filter bubbles make this difficult, when one emotional perspective enforces engagement with evidence that supports one's viewpoint rather than also looking at other sides. On the other hand, social media also provide the opportunity to find posts that show other sides. This is not to say that there is this complexity in every situation. In some situations, people do things that are simply morally wrong without further qualifications. However, things are complex in many situations, and we should first be open to other insights, narratives, and experiences before making up our minds. This is what can happen in emotional deliberation, and it can also happen online.

There is another important dimension in which emotions can play a crucial role in our moral deliberation, and that concerns the role of moral exemplars. Linda Zagzebski (2021) has argued for the importance of narratives of moral exemplars. These concern people whom we would like to imitate and this is grounded in our emotions of admiration and in aligning ourselves with the emotional responses of others. She describes a moral exemplar as a "paradigmatically good person […] who is most admirable […] A person who is admirable in some respect is imitable in that respect. The feeling of admiration is a kind of attraction that carries the impetus to imitate with it" (Zagzebski, 2013, p. 201). Moral exemplars can thus inspire people to practice the virtues that the former possess. Crucial to the functioning of moral exemplars as promoters of virtue is the emotion of admiration since it is through this emotion that these exemplars are identified in the first place and that people are led to imitate them. Zagzebski considers admiration as having a cognitive element since the emotion allows the experiencer to evaluate the object of the emotion as precisely admirable. In other words, moral exemplars also (implicitly) utilize the notion of cognitive emotions, particularly that of admiration.

In the following section, we will look at Confucian philosophy, a tradition that also includes a cognitive understanding of emotion and thus has ideas that resonate with affectual intuitionism and emotional deliberation. More importantly, Confucian philosophy provides an ethical dimension for emotions, which we believe can

strengthen our argument above that emotions can be used for ethically positive goals like mitigating OP instead of being ignored or eliminated. Moreover, Confucian philosophy has ideas on how to utilize emotions to orient people towards ethical ends. These ideas, in turn, can serve as inspiration for methods of reducing OP that precisely capitalize on human emotions rather than dismiss them.

4.5. Confucian Philosophy and Cognitive Emotions

Confucian philosophy is a promising conceptual resource for formulating and developing emotion-based approaches to overcome OP. This is because the idea of cognitive emotions and strategies for utilizing cognitive emotions for ethical ends can be implicitly found in Confucian philosophy. Furthermore, Confucian ethics emphasizes respect and interpersonal harmony, which are values that can be useful in mitigating OP. In this section, we first describe cognitive emotions' integral role in the Confucian view of human existence and flourishing. We then discuss methods in Confucian philosophy for utilizing cognitive emotions for ethical purposes, which in turn can contribute to emotion-addressing approaches to OP. These methods are: extension (of emotions), reverencing human nature, and using moral exemplars.

4.5.1. The Confucian View of Human Existence, Flourishing, and Cognitive Emotions

Confucian philosophy considers humans as emotional creatures by virtue of their nature and embeddedness in the world. The Confucian classic, *Doctrine of the Mean* (*Zhongyong* 中庸), describes humans as having natures that consistently respond with emotions to the things and affairs they encounter (Nelson, 2018). These emotions, however, are not wholly "unintelligent." Instead, they are inextricably intertwined with human cognition. This is represented by the Confucian concept of the "heart-mind" (*xin* 心). Although the character of the heart-mind illustrates the physical heart, it is much more than just a bodily organ in Confucian philosophy. Neither is it merely the seat of one's affect, as the conventional understanding of "heart" today might connote. Rather, the heart-mind is the "locus, or faculty, of emotions, cognition, evaluation, and judgment within the self" (Virag, 2017, pp. 109-110). The heart-mind represents the Confucian idea that there is no sharp distinction between a person's cognitive and affective capacities and that these capacities and the cognitive emotions they produce are central to the human being.

The heart-mind's connection with cognitive emotions is seen in the thought of the Confucian philosopher Mencius (372-289 B.C.E). Mencius' idea of the "four sprouts" closely corresponds with the idea of cognitive emotions since both ideas consider emotions as a form of perceiving value. For Mencius, the heart-mind contains four "sprouts" or natural emotions: compassion, shame, deference, and approval and disapproval. These emotions have a cognitive character since experiencing them involves the evaluation of the object, feature, or situation at which they are directed. An example of this can be inferred from Mencius' example of a child falling into a well. He says that anyone who witnesses the child would have "a feeling of alarm and compassion — not because one sought to get in good with the child's parents, not because one wanted fame among their neighbors and friends, and not because one would dislike the sound of the child's cries" (*Mengzi* 2A6. 3-4/ Mencius, 2008, p. 46). In this instance, the feeling of compassion is directed toward a feature of the situation in which it arises, namely, the child and its imminent endangerment. The feeling makes this feature salient to the experiencer of compassion (Hu, 2019). However, this feature is implied to be intimately related to a value — not those involved with personal gain but the child's well-being (Liu, 2002; Hutton, 2002). Through the emotion of compassion, the experiencer tacitly appraises the child's impending suffering as a threat to a value (such as the child's well-being) and (possibly) as a reason for acting (e.g., helping the child) (Wong, 1991).[80] This Mencian understanding of compassion thus resonates with the idea of affective intuitionism mentioned above.

The heart-mind and its cognitive emotions also represent human moral potential. By developing these emotions into virtues, humans actualize this potential and attain flourishing. The Confucian vision of flourishing can be considered individually and collectively. Individually, flourishing involves possessing moral virtues such as benevolence and righteousness (Kim, 2020). People should ideally attain mastery of these virtues in such a way that enacting them comes effortlessly. Moreover, Confucians believe that manifesting these virtues could inspire others towards goodness. The virtue of respect for other persons is particularly prized by Confucians, being considered an essential component in the *junzi* (君子), or virtuous person's repertoire of virtues (e.g., *Analects* 12.5/ Confucius, 2003, p. 127). On the other hand, at the collective level, Confucian flourishing consists of a morally virtuous community

[80] Scholars of Mencius' philosophy have spilled much ink regarding how exactly the four sprouts or emotions are connected to cognition. For a modest summary of these scholars' views, see Hu (2019).

characterized by benevolent political institutions, respectful and harmonious social relationships, and sufficient material well-being for its members.[81]

Cognitive emotions are thus an integral component of Confucian flourishing. Because of this, Confucians do not seek to suppress or eliminate them even if they prove potentially detrimental to flourishing. Rather, Confucians believe that cognitive emotions should be educated so they can be expressed in a morally appropriate manner. In at least Mencius' version of Confucianism, cognitive emotions play a more explicitly central role in flourishing since it is precisely through their ethical cultivation — that is, the cultivation of the four sprouts — that people acquire the moral virtues most central to flourishing. To wit, the cognitive emotion of compassion (*ceyin* 惻隱) can be cultivated into the virtue of benevolence (*ren* 仁), the cognitive emotion of shame (*xiuwu* 羞惡) can be cultivated into the virtue of righteousness (*yi* 義) the cognitive emotion of deference (*cirang* 辭讓) can be cultivated into the virtue of propriety (*li* 禮), and the cognitive emotion of approval and disapproval (*shifei* 是非) can be cultivated into the virtue of wisdom (*zhi* 智).

In summary, the relationship between the cognitive and the affective, as conceived in Confucian philosophy, resonates with contemporary cognitive theories of emotions discussed in section 4.4. What makes the Confucian — particularly the "Mencian" — conception of cognitive emotions distinctive is the idea that such emotions have a constitutive role in the normative life and flourishing of human beings.[82] Cognitive emotions are the incipient material that can be cultivated into virtues, central to the Confucian vision of flourishing. Unsurprisingly, the Confucian tradition has developed distinctive methods for precisely developing these virtues. Put alternatively, these methods help orient cognitive emotions toward ethical ends. We believe these methods are worth exploring since they can serve as inspiration for ways of mitigating OP that likewise use emotions instead of dismissing them. We describe these methods below.

---

81 Harmony is an especially valued quality of collective flourishing. Confucian harmony does not entail uniformity or sameness. Rather, it requires diverse elements respectfully working together for the common good (Li 2014). Confucians deemed this diversity as something that could make society more existentially rich and functionally effective than the case where all of its members simply conformed to exactly the same standards.

82 We say "Mencian" conception since there are strands of Confucian thought that do not follow Mencius. In particular, the Confucian philosopher Xunzi (310-220 B.C.E.) criticizes Mencius' view of human nature, claiming instead that humans have bad natures that are devoid of the four sprouts or any such moral direction.

4.5.2. Mencian Extension

"Extension"(*tui* 推) refers to the method of "extending" the abovementioned four sprouts of compassion, shame, deference, and approval and disapproval. Extension can be considered a kind of emotional deliberation. It requires reflecting on and simulating one's previous moral feelings (i.e., one of the sprouts) to motivate oneself to apply those same feelings to a situation or set of objects to which one has not yet applied them (Mencius, 2008). The specifics of how extension works are a matter of debate for scholars.[83] However, one helpful way of understanding extension, suggested by philosopher David Wong, is in terms of "reinstatement" (2015, 2023). This is the idea that a certain situation can arouse an emotion if it resembles a previous situation in which one has also felt the same emotion. The resemblance need not be relevant, nor the comparison between the two situations be conscious. For instance, one can unconsciously feel anxiety towards an officemate simply because the latter wears the same checkered shirt that one's school bully wore in the past. Wong, however, argues that Mencius' extension involves a conscious form of reinstatement. To see this more clearly, let us take a look at the most detailed account of (an attempt at) extension. This is Mencius' dialogue with King Xuan, where the former tries to guide the latter into extending his compassion (*Mengzi* 1A7.3–7.13/ Mencius, 2008, pp. 7–11).

The dialogue begins with the king asking Mencius if he has what it takes to care for his people (who are currently suffering from his misrule). Mencius then reminds the king of an incident where the latter spared an ox about to be sacrificed for a ritual because he saw that it resembled an innocent person about to be executed. The king confirms that this happened, after which Mencius states that this incident reveals that the king has, indeed, a heart sufficient to care for the people. This is because, according to Mencius, the king spared the ox not out of stinginess but out of compassion. This makes the king realize that he is not actually sure why he spared the ox, which prompts him to reflect back on this experience. After a brief moment, he comes to the realization that it was indeed out of compassion that he prevented the ox's death because it looked like an innocent about to be killed. Realizing that his heart has compassion, the king then asks Mencius how this could make him care for the people. Mencius responds by saying, in essence, that all the king needs to do is to extend compassion to his people. Mencius eggs the king on by using analogies that emphasize how incongruous it is that the latter can show compassion to an ox but not

[83] For an anthology containing different views regarding this matter, see Liu & Ivanhoe (2002).

his people. For instance, he compares it "to someone who claims to be able to lift 500 pounds but not a feather, and also to someone who claims to see the tip of an autumn hair but not a wagon of firewood" (Sta. Maria, 2020, p. 62). Mencius then again juxtaposes the king's compassion for the ox with his lack of compassion towards his people. He says, "In the present case your kindness is sufficient to reach animals, but the effects do not reach the commoners. Why is this case alone different?" (*Mengzi* 1A7.12/ Mencius, 2008, p. 11).

The above dialogue can be understood as Mencius' attempt to reinstate the king's compassion for the ox. However, the new object of the reinstated emotion would be the king's suffering people. Mencius does this by first prompting the king to recall and reflect on the past experience where he compassionately spared the ox. He then guides the king to find relevant similarities or associations between the cases of the innocent man (whom the king thought the ox resembled), the ox, and the king's people (Wong, 2017). This is not a purely cognitive exercise because in leading the king to remember the past instance of compassion, Mencius is also making him re-live or re-experience the aforementioned feeling. Thus, the entire exercise of finding similarities is "imbued" with the feeling of compassion that the king felt before. This arguably helps the king establish relevant similarities between the cases because, in reexperiencing his past compassion, the values and value-laden features to which that feeling was directed are made more salient to him. This, in turn, can facilitate the identification of any resemblances between those values and features, and those in the new situation to which his compassion is now supposed to be directed. To put it more concretely, by reliving the compassion he felt for the ox, the king might realize (in a cognitive cum emotional way) that what appeared most salient for him during that occasion when he was experiencing compassion was the innocence and vulnerability of the animal. When he directs his attention to his suffering people while still having the re-experience of compassion fresh in his heart-mind, he might also realize that his people deserve compassion because of their innocence and vulnerability. Ultimately, this can motivate him to show or reinstate compassion for his people.

Mencius introduces another example of extension in his dialogue with the king. Mencius says, "Treat your elders as elders, and extend it to the elders of others; treat your young ones as young ones, and extend it to the young ones of others" (*Mengzi* 1A7.12/ Mencius, 2008, p. 11). To put this generally, Mencius recommends extending one's feeling of compassion or care from one's intimates (or ingroup) towards non-familiar people (or the outgroup). In this case, the emotion is extended, not from a specific past instance but from a person's general experience of relating to

their loved ones. Extending care from one's intimates towards non-intimates fits the Confucian idea of working with emotions rather than against them. This is because extension does not seek to eliminate people's natural affection for their family or ingroup. It even capitalizes on such natural affection by using it as a model or frame to motivate people to care for the outgroup. Although Mencius does not describe this kind of extension in detail, we can surmise that it also involves reinstating the emotion towards the new target by attempting to establish resonances between the previous and new objects of the emotion while simulating the experience of the emotion itself.

Extension, then, can inspire recommendations to mitigate OP in a way that works with users' emotions instead of against them. The specifics of these recommendations will be discussed in section 4.6.1. For now, we shall move on to the next Confucian method of utilizing cognitive emotions for ethical ends.[84]

4.5.3. Reverencing Human Nature

At this point, we would like to introduce a notion we call 'reverencing human nature.' By this, we mean the act of giving serious, respectful, and reflective consideration to an idea of human nature. This act can then motivate one to adjust one's behavior according to the ethical implications of the idea of human nature. Our notion of reverencing human nature combines two different Confucian ideas. One is the idea of 'reverential attention' by the Neo-Confucian philosopher Zhu Xi (1126-1271 C.E.). The other is Confucian ideas of human nature. In other words, reverencing human nature is the method of paying reverential attention to Confucian (or Confucian-inspired) conceptions of human nature. We will discuss the idea of reverential attention and the Confucian conceptions of human nature consecutively in what follows.

"Reverential attention (*jing* 敬)" refers to an attitude that Zhu Xi considered the foundation of cultivating moral virtue. According to scholars Stephen Angle and Justin Tiwald, reverential attention consists of doing 'two things at once: 'first, we are

[84] It might be asked how people will know when to (or to whom to) apply extension. Perhaps there are cases when extension should not be done, such as when the proposed objects of extension are people who have highly immoral positions or values. We do not have space to delve thoroughly into this matter. We hope the following response will suffice for now: We intend to translate the idea of extension as a digital prompt that will appear while users browse the internet (more of this will be discussed in section 4.6.1). Now, that prompt can be designed to not "activate" when the situation does not call for it. As to what situations might not call for extension, we recommend the reader to look at Read (2022), which might (partly) answer this question.

to concentrate on a task or object of inquiry with a certain kind of respect, which – second – undercuts selfish intentions, including ones that are easy to overlook or misapprehend' (2017, p. 148). Although Angle and Tiwald do not explicitly describe it as such, reverential attention can be understood as a cognitive emotion. This is hinted at by Zhu Xi's equating of reverential attention with the emotion of awe (Zhu Xi 2019). Awe can be considered a cognitive emotion because it involves an evaluation or appraisal "of there being something normatively important that lies beyond oneself" (Angle, 2009, p. 86). Moreover, it is not uncommon for awe to frame this normatively important object as more significant than one's self-serving interests. This, then, prompts the experiencer of awe or reverential attention to undercut their selfish intentions, as stated above. The objects of reverential attention can range from tangible things, such as people, to "intangible" entities, like concepts. This may sound esoteric or foreign to everyday life. However, the underlying idea is that reverential attention is simply the act of respectfully and seriously considering some object, whether concrete or abstract.

This leads to the second idea that constitutes "reverencing human nature," namely, Confucian conceptions of human nature. We would like to use two distinct ideas of human nature from Confucian philosophy. The first is Mencius' idea of human nature. For Mencius, human nature is good because it contains the four sprouts (as discussed in 4.5.1.), which are natural and morally-oriented emotions. We believe that directing reverential attention to the Mencian concept of human nature can encourage prosocial and ethical attitudes. First, it can help users value the sprouts — that is to say, the moral potentialities — of other people, thereby encouraging users to refrain from influencing people with behaviors that could "damage" these sprouts. For instance, one of the sprouts is the emotion of deference or respect, which can develop into the virtue of propriety. Paying reverential attention to this sprout could motivate users to avoid becoming an example of disrespectful behavior, which might influence others against cultivating their own sprout of respect.[85] Paying reverential attention to the goodness of human nature might also encourage users to have a charitable view of others, particularly those who belong to the partisan outgroup. This is because users would be led to trust that, despite differences in opinion and, indeed, despite the bad beliefs and behaviors of their opponents, the latter are not likely to be irredeemable "moral monsters." Rather, they are still beings with natural inclinations towards goodness (though this inclination might have been frustrated by these peoples'

[85] As we shall show in the following section, this idea connects to the Confucian notion of "influence modelling" which emphasizes the importance of being a good moral example for others.

badness). One could object that believing in the natural goodness of humans is naïve; some people appear to incorrigibly embrace morally despicable views and some even practice them. However, the idea of good human nature is not meant to deny the existence of such people. Instead, it is meant to remedy the tendency of users to demonize the outgroup — that is to say, to prematurely impute to them an irredeemable level of malice — by making users open to the possibility that the outgroup is otherwise than what they think. However, if the outgroup proves to have reprehensible views and is obstinate in adhering to them, then users would be right to refuse to treat them amicably.

The other concept of human nature we choose is that of the Confucian Dai Zhen. Dai Zhen claims that all human beings have natural desires for the goods that they require as the living beings that they are (Tiwald 2010). These are desires for food and drink, love, shelter from the elements, "freedom from persistent anxiety, and lifelong companionship" (Tiwald 2010a, p. 88, n.14). Paying reverential attention to this idea of universal human desires for likewise universal human needs can help users develop a greater sense of sympathy with others because it can help them realize that they have commonalities with every person they encounter despite differences in group loyalties. Focusing on what people share in common, in turn, can lessen political polarization (Mutz, 2006; Sunstein, 2017).

To wrap up, our proposed notion of reverencing human nature can be understood as a kind of emotional deliberation on Confucian conceptions of human nature, which can motivate prosocial or moral attitudes towards the outgroup and, thereby, reduce polarization. The recommendations that can be drawn from reverencing human nature will be discussed in section 4.6.2. For now, we go to the last Confucian method of using cognitive emotions for ethical ends, namely, moral exemplars.

4.5.4. Moral Exemplars

In section 4.4., we mentioned that the philosopher Linda Zagzebski has discussed the role of cognitive emotions in relation to moral exemplars. The Confucian tradition has insights that can increase the effectiveness of moral exemplars for ethical development. We discuss two of these insights below.

The first is that moral exemplars can be "imperfect." Confucianism has used such imperfect exemplars to emotionally engage and motivate aspirants to walk the moral path. The fallibility of Confucian moral exemplars can be understood in two ways. One is that there can be an exemplar who possesses most if not all of the virtues — what Confucian scholar Amy Olberding calls a "total exemplar," — yet nevertheless

illustrates flaws or idiosyncrasies (2012). An example of this is Confucius, as portrayed in the *Analects*. Although held up as one who has mastered virtue, Confucius is also shown to have several foibles, such as being prone to emotional outbursts and cutting remarks, as well as having a unique and sometimes misunderstood sense of humor. The fallibilities of these total exemplars humanize them and make them and the ideal they represent more emotionally relatable to aspirants. This can encourage aspirants to strive towards such ideals since they are reminded that even the exemplars that embody them are but fallible humans.

Another imperfect exemplar is what Olberding calls a "partial exemplar" (2012). These exemplars excel only in one or a few virtues but still need work to become virtuous in other respects. Partial exemplars can inspire aspirants to focus on the more modest goal of attaining a particular set of virtues, such as the one that the partial exemplar possesses or lacks, instead of despairing over the impossibility of emulating the total exemplar and their complete repertoire of virtues. Moreover, since partial exemplars only excel in a few virtues and therefore might have some of the typical "vices" that people have, they can even prove more "human" or relatable to aspirants than total exemplars are, and thereby be more effective in motivating aspirants. In the *Analects*, an example of a partial exemplar is Confucius' disciple, Zigong (Olberding, 2012). Zigong is portrayed as both eager and adept at learning "ritual," or those socio-ethical norms Confucius believed were vital to society's well-being. However, his excessive appreciation of ritual often leads to rigidity and harshness in evaluating people who, in his eyes, fail to live up to ritual standards. Thus, Confucius must continually remind him to be more flexible and sympathetic towards others. Aspirants would likely be able to emotionally relate to Zigong, seeing themselves in him and finding comfort and inspiration from the idea that even someone as "flawed" as him is considered one of Confucius' most notable followers.

The second Confucian insight that can improve the use of moral exemplars has to do with their understanding of role modeling. According to scholar Cheryl Cottine, Confucians emphasized two dimensions of role modeling (2016). First is what she calls "imitation modeling," which emphasizes the need to imitate moral exemplars. The second is "influence modeling" which focuses on the importance of *being an exemplar* or model for others. For Cottine, these two dimensions are inseparable components of Confucian moral development. In other words, Confucian aspirants were aware that to grow in virtue not only meant imitating exemplars but also becoming a good model for others since their own actions invariably influenced their fellow human beings.

As we will demonstrate in section 4.6.3 below, these two insights can be incorporated into using moral exemplars in the online realm. Similar to the previous methods, our idea of using moral exemplars will involve a kind of emotional deliberation on the exemplar, which in turn can encourage users to develop attitudes and behaviors that can mitigate OP.

4.6. Recommendations: Using Confucian ideas on emotions to help overcome OP

How can the Confucian ideas discussed above be applied to online interactions? Two challenges hamper this application. First, these Confucian ideas originally assume that their users are situated in an environment unmediated by digital technologies. The kind of human relations and interactions afforded by online platforms are quite distinct from those of their physical counterparts. These differences can affect the forms that these Confucian ideas might take. Furthermore, the contemporary, pluralist nature of society (both offline and online) might not be wholly compatible with the historical versions of these methods since the latter were employed in a non-pluralist context. Thus, in our attempts below to translate these Confucian ideas into the online setting, we will take note of these considerations and modify the aforesaid ideas accordingly. We also wish to stress that the effectiveness of these (modified) ideas is not set in stone and should be subject to empirical inquiry and ethical evaluation. As philosophers, our area of competence lies with articulating theories that would need to be translated into concrete applications by experts in other fields, such as, in this case, designers of digital architecture. Therefore, we present these recommendations as conceptual points of departure for the appropriate experts to build on.

4.6.1. Applying "Mencian Extension"

Applying Mencius' method of extension to the online realm can take inspiration from Saveski et al.'s idea of an "empathic prompt" (2022). In their experiment, Saveski et al. exposed online users to tweets from people of the opposite political spectrum. Before doing so, however, they gave the experimental group a reminder (i.e., the empathic prompt) stating, "If you see views that differ from your own, think of a time when you and a friend had a difference of opinion" (888, 2022). They found that group members were more likely to understand why people of the outgroup held the views they did.

A Mencian-inspired version of an empathic prompt would request users, before interacting with the outgroup and their content, to think of their care for their loved ones and then try to apply such care to the outgroup. This aligns with Mencius' idea of extending one's care for one's intimates towards non-intimates. Another way a Mencian empathic prompt could work is by telling users to recall an instance where they felt care or compassion, encouraging them to even simulate and feel those same emotions once more. The prompt then advises users to apply these emotions when encountering the outgroup. This second way takes inspiration from Mencius' attempt to make King Xuan extend his compassion for his people by leading him to reflect on his compassion for the ox.

A feature of Mencian empathic prompts is that they would actively lead users to imagine and reflect on the occasions when their original emotion of care manifested. The idea of imagining and reflecting is, again, in line with the method of extension applied by Mencius to King Xuan, where the latter, in recalling his compassion for the ox, is implied to have reimagined and reflected on the past incident. The Mencian prompt would make users pause in order to actively imagine and reflect on their experiences. To ensure that this pause happens, the prompt could include a design feature that prevents users from interacting with the webpage for a brief period of time. This is similar to apps like Eye Saver, which urge users to rest their eyes from viewing the computer screen by inhibiting computer use for some seconds. One could object that compelling users to pause would be too intrusive and could trigger emotions that are actually conducive to polarization, such as aggression. We have two responses to this, which we do not pretend to be able to wholly address the issue. First, like apps such as Eye Saver, the Mencian empathic prompt could allow users to cancel the pause if they feel it is too restricting. Second, the Mencian empathic prompt, including the pause it imposes, can be activated only occasionally so as not to pester users.

To partially support the effectiveness of these ideas we have just mentioned, we would like to reiterate that Saveski et al.'s empathic prompt has been efficacious in promoting empathic attitudes in users. Nevertheless, there is a salient difference between Saveski et al.'s prompt, and the Mencian one; whereas the former is simply a written reminder for users, the latter would be a period of time where users could pause and actively use their imaginative and reflective capacities, as well as their (cognitive) emotions, to extend feelings of care towards the outgroup. Because of this difference, the result of Saveski et al.'s empathic prompt cannot fully prove that our ideas will also be effective. However, the Mencian prompts can be seen as a

distinctively Confucian way of building upon the implications of Saveski et al.'s findings and possibly improving their strategy for engendering empathy.

4.6.2. Applying "Reverencing Human Nature"

The method of reverencing human nature can be translated to the online context similar to Mencian extension. That is to say, it can be a prompt where users are led to reflect on an object while simulating an emotion at the same time. In the case of this "human nature prompt," the object would be the aforementioned Confucian conceptions of human nature. The prompt would ask the users to pay reverential attention to these conceptions, which, in turn, involves simulating the feeling of respect or awe towards them. There can be two forms of this prompt based on the two Confucian concepts of human nature mentioned in section 4.5.3. The first form has for its object the Mencian concept of good human nature. This prompt could lead users to pay reverential attention to the natural goodness of human beings, thereby motivating them to interact with others in a way that nourishes others' four sprouts or moral potential. The prompt could also motivate users to treat others, particularly the partisan outgroup, charitably, since it makes them appreciate the idea that all humans have naturally good inclinations. Like the Mencian empathic prompt, this prompt could be activated before users are exposed to the outgroup. Likewise, the prompt can lead users to set some time for paying reverential attention to the concept of human nature by temporarily inhibiting interaction with the web browser, or, if that is too restrictive, by simply providing reminders that encourage such reflections (these could be options that users could choose for themselves). The second form of the human nature prompt has for its object the Dai Zhenian concept of human nature with its focus on natural human desires and needs. This prompt can work like the previous form, except that it will lead users to pay reverential attention to the common needs of humanity as outlined by the aforesaid idea of human nature, thereby potentially helping users develop sympathy with other people.

Another way in which the Dai Zhenian concept of human nature can be translated online, particularly on social media platforms, is through post flairs. Post flairs, like those in Reddit, are colored labels that can be attached to posts to let people know what a post's topic is about. Examples of these would be "Politics," "Sports," and "Culture." Applying the Dai Zhenian concept of human nature would mean categorizing the posts of users based on the needs that Dai Zhen claims all humans desire. For instance, a post about welfare programs (whether or not the posts support such programs) can be categorized under the flair "freedom from consistent anxiety."

Categorizing flairs in this way can remind users that even the unpalatable posts of outgroup members might ultimately be rooted in the natural desires and goods common to all humans. This can engender sympathy for outgroup members, especially if combined with the aforementioned Dai Zhenian prompt.

An objection to using the method of reverencing human nature is that it is inappropriate for our contemporary pluralistic setting due to the potentially exclusionary definition of humanity that the concept of "human nature" implies. There is no space in this paper to reply to this objection comprehensively. Our short response is that the method of reverencing human nature need not adapt an essentialist and exclusionary idea of human nature for it to be effective. The method can simply take inspiration from Mencius and Dai Zhen's conceptions of human nature to make a modest account of a human being's possible moral potentialities and natural needs. This modest account would refrain from making thick, metaphysical, and absolutist claims on what defines a human. For instance, whereas Mencius' idea of human nature claims that having the feeling of compassion is a condition for being human, this modest account would simply state that compassion is a general moral potential humans have and that developing it is desirable for individual and collective flourishing. Likewise, Dai Zhen's conception of human nature would be translated as the simple claim that humans generally have desires for a common set of needs, which need to be fulfilled for flourishing. The idea of common requirements for human flourishing would not be outdated because modern theories, such as objective list theories of well-being, are based on this idea.[86,87]

4.6.3. Applying the Use of Moral Exemplars

Using moral exemplars in the contemporary online context can be challenging due to several factors. We divide these challenges into two kinds, namely, the epistemic and

---

[86] An example of such an objective list theory is the one articulated by philosopher Richard Kraut (2007), who uses the concept of human nature to establish what is good for humans generally. Another arguable example is Martha Nussbaum's capabilities approach (2011), which claims that exercising and developing certain human capabilities is objectively and good for all humans.

[87] Another objection to the idea of the reverencing human nature prompt is that which was mentioned against the Mencian prompts, i.e., it might be too intrusive and might actually cause negative attitudes in users. Our response to this is the same we give to defend the Mencian prompts. However, we would like to reiterate as well that our responses are not meant to be foolproof and that our recommendations are only meant to be points of departure for more refined ways of applying our Confucian ideas to the online realm. Again, we leave the formulation of these more refined ways to experts in digital architecture

ethical. The epistemic challenges are as follows. First, admiration of moral exemplars requires that much about the latter should be known, such as detailed accounts of their virtuous lives and meritorious deeds. However, the amount of information users post about themselves online usually does not equal this knowledge of exemplars. This is accentuated by the fact that users, particularly of populated social media platforms like Facebook, often encounter people they are not familiar with and have little knowledge of (Mossner & Walter, 2024). Moreover, the personal information that people present online can be highly inaccurate, making it difficult to ascertain if they are truly worthy of admiration.[88] On the other hand, the ethical challenge is this: moral exemplars, such as those in Confucianism, might embody thick moral principles valued only in the ethical tradition to which they belong. This can be difficult to square with the pluralistic nature of current society, including the online realm, where people adhere to a multiplicity of possibly incommensurable values.

We can partially address the epistemic challenges using the Confucian idea of imperfect exemplars. This idea shows that an exemplar can still fulfill its function even if it excels only in one or a few virtues. In the context of the online environment, this means that even if a user's knowledge of another person is limited or even inaccurate, such that they only know of the good behavior they do, that might still suffice to make that person an imperfect exemplar for the user. Utilizing imperfect exemplars can give users the impression that adopting behaviors that lessen OP – a task that might require them to go against their own parochial impulses – is achievable since even flawed people can do it. Considering this, a concrete way to translate moral exemplars online can be the employment of badges and comment ranking. Badges, such as those found on Facebook, are labels attached to users that inform others of their contribution, usually related to a group. For example, a user who posts or replies frequently on a group page can acquire the "Top Fan" or "Top Commenter" badge, which indicates that they are a lively participant in the group. Perhaps there can be a badge then that attaches to users who interact with polarizing people or posts respectfully and in a manner that lessens hostilities. The badge could be titled "responsible commenter" and could be designed in a way that catches the attention of users. This feature could

[88] One could object that a moral exemplar can be fictitious. While this is possible, we believe that, in the online context, perhaps having a fictitious moral exemplar (say a "fake" account run by a bot or AI) might not be effective since users might not be encouraged to think that the virtues the exemplar displays are achievable by humans. Moreover, if the fictitious exemplar pretends to be a real person, then the discovery that it is indeed fictitious might also discourage users and even push them towards an opposite, vicious, path.

go hand in hand with comment ranking. Comment ranking is a feature in social media, such as Facebook and Youtube, where the highest-ranking comment appears as the first comment under a post. It is usually already visible to users even if the latter has not yet chosen the option to view all the comments on the post. It might be good then to have a feature that makes the most respectful comment, or the comment most conducive to harmonious interactions, be the first comment under a post. This comment could be given by a user who already has a badge for responsible commenting. Such a user, being immediately and particularly visible to others due to their badge and the comment ranking system, might thus serve as an imperfect exemplar for other users when it comes to respectful and harmonious interactions. Again, other users might not necessarily know everything about this exemplar. However, witnessing the exemplar's good attitude and behavior might be enough to evoke admiration from them, inspiring them to imitate the former.[89]

Another way to address the epistemic challenges is to utilize the Confucian idea of influence modeling. Again, influence modeling entails focusing on the idea that one can also influence the behavior of others, so one should strive for ethical improvement. In the online context, this means that users should be urged to act as examples for fellow users. This addresses the epistemic challenge of limited knowledge since the users have to focus on themselves and not on others who are anonymous. As a design feature, this idea can be translated as a prompt that works similarly to the empathic prompts discussed above in the sense that it urges users to spend some time directing and experiencing a cognitive emotion towards a certain object. Specifically, this prompt can lead users to pay reverential attention or admiration to their own moral potential or potentially perfected moral selves, thereby motivating them to fulfill such potential by acting in virtuous ways that are opposed to those behaviors that increase polarization. The idea of influence modeling can also be applied to the online realm by simply being an ethical principle that users attempt to enact in their interactions with other users. In other words, users will strive to act appropriately towards other netizens to influence the latter positively and not in a way that will lead them to vicious, polarizing, behavior. Practicing influence modeling seems especially appropriate for so-called social media "influencers" because of their ability to sway

[89] A question that arises with this whole idea of providing badges and ranking comments is who exactly decides which user gets a badge and/or which comment is ranked first? Our tentative answer is that the AI algorithm can handle this task. However, we are keenly aware that this opens up a whole host of other questions and problems. As such, we defer to the experts in digital/AI design (and perhaps digital sociology) to provide a better solution to this issue.

large groups of netizens and, consequently, their considerable potential to heighten or lessen polarization.

Finally, when it comes to the ethical challenge, this can be addressed by making the exemplars embody moral attitudes or values that are thin and common enough to be agreed upon by most, if not all, online users. Put alternatively, the criteria for picking an exemplar would revolve more on whether a person refrains from acting in disrespectful or, precisely, polarizing ways that online users would generally find disagreeable than whether the person lives up to a thick and specific account of a particular moral tradition. Thus, for instance, an online user would merit the "responsible commenter" badge more likely because they refuse to engage in abusive or incendiary interactions with other users, rather than because they are fulfilling all the conditions of, say, the historical understanding of the Confucian virtue of benevolence.

Last but not least, we would like to highlight that the Confucian idea of the moral exemplar could also be a guiding light for owners of social media platforms. Through a Confucian lens, we can argue that those who hold leadership positions are only deserving of these if they are moral exemplars. Current developments are a far cry from this. It probably also does not help that many leading 'tech bros' find their philosophical inspiration in frameworks that are ethically dubious, and which seem to serve to legitimize immoral behavior rather than providing for a critical lens.[90] As long as leading politicians and media from around the world and the whole political spectrum use X as the ultimate platform for exchanging information, the role of the owner of this platform should be to live up to the highest moral standards. Unfortunately, the reality seems to be quite the contrary.[91] Here as well, Confucianism can provide a critical source of inspiration. Just as Mencius provided kings with critical feedback and appealed to demanding standards of virtues such as compassion and

---

90 See, for instance: Ord, T. (2023, October 11). Why longtermism is the world's most dangerous secular credo. Aeon. https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo, and Burkeman, O. (2022, October 31). Elon Musk's idea of 'free speech' will be bad for Twitter – and for all of us. The Guardian. https://www.theguardian.com/commentisfree/2022/oct/31/elon-musk-twitter-trump-tesla-longtermism

91 For a few recent examples concerning problematic aspects of Musk's ideological outlook, eg facilitating Trump, hate speech and antisemitism, see: Times of India. (2023, October 14). Elon Musk said 'I am Dark MAGA'. What exactly is Dark MAGA? The Times of India. https://timesofindia.indiatimes.com/world/us/elon-musk-said-i-am-dark-maga-what-exactly-is-dark-maga/articleshow/113989846.cms, and CNN. (2023, November 17). Elon Musk reveals his 'actual truth'. CNN. https://www.cnn.com/2023/11/17/business/elon-musk-reveals-his-actual-truth/index.html.

care, the role of contemporary philosophers advising tech developers should be equally critical and demanding, and technology leaders should be aware of the enormous moral responsibilities they carry.

4.7. Conclusion

In this paper, we discussed existing approaches to analyzing and overcoming online polarization (OP). We discussed how mainstream approaches fall prey to a rationalistic bias, overlooking the role of emotions. We then discussed an alternative approach to OP that emphasizes emotions, namely, the theory of affective polarization. However, this theory is in danger of framing emotions as simply the fuel that drives polarization, thereby relaying the idea that emotions should be simply dismissed or eliminated. In contrast to this, we claimed that it is more just and effective to work with emotions than against them when it comes to mitigating OP. We supported this by using insights from cognitive theories of emotions. In particular, we discussed the following crucial features of cognitive emotions that help justify the idea that such emotions should be appreciated rather than eliminated: their intrinsic cognitive aspects and rationality, the capacity of emotional deliberation, which is grounded in care and sympathy, and the role of moral exemplars. To further support our claim that utilizing emotions is a viable way of remedying polarization, we turned to Confucian philosophy which highlights the integral role of cognitive emotions in ethical flourishing and, more importantly, possesses methods (i.e. extension, reverencing human nature, and using moral exemplars) for cultivating cognitive emotions towards ethical ends. Drawing inspiration from these ideas, we recommended possible strategies for lessening OP that capitalize on natural human emotions instead of negating them. Ultimately, we hope that our paper can make a modest contribution to addressing the problem of OP and perhaps even to the polarization that occurs outside the digital space, especially because both online and offline polarization have become particularly grave impediments to the harmonious coexistence of people in this day and age.

Interlude 4

The previous chapter demonstrated how Confucian philosophy can serve as a conceptual resource for thinking about and harnessing cognitive emotions in a way that may potentially reduce political polarization on social media. In doing so, the previous chapter has shown that Confucian philosophy can contribute to the pursuit of the digital good life at the political sphere. The following chapter shall go beyond the sphere of the political and proceed to the level of the "world" by examining AI systems. This kind of technology has a global reach, not only in the sense that its effects are felt worldwide, but also because such effects are present in virtually all aspects of human life. Due to their deeply pervasive and influential presence, AI systems have the great potential for conceptual disruption, that is, the ability to challenge the meanings and functions of concepts by which we understand and navigate the world. Although conceptual disruption is not necessarily bad, it can nevertheless hinder discourse on AI ethics and governance, as the influence of AI itself can disrupt the very concepts involved in this discourse. Thus, it might be necessary to create measures that mitigate or even eliminate the effects of conceptual disruption. The following chapter will explore how Confucian philosophy can contribute to this endeavor by demonstrating how its ideas can help engineer a concept with conceptual resilience, or the ability to remain immune to conceptual disruption, by introducing the notion of an "indicative concept." The chapter will also demonstrate how this Confucian-inspired concept can be applied in a specific discourse in AI ethics – namely, the debate on whether AI can be considered a friend to humans – in a manner that potentially promotes the flourishing of the stakeholders involved. By presenting all these, the chapter aims to highlight how Confucian philosophy can contribute to the pursuit of the digital good life at the level of the world. This is not only because the digital technology involved, i.e., AI systems, is a global phenomenon, but also because concepts are those by which people fundamentally understand and navigate the world.

Following the next chapter, the dissertation will end with the conclusion, which will answer the main research question. The conclusion will provide a summary of the findings from chapters 2-5, identifying the *specific insights* that Confucian philosophy can offer regarding the pursuit of the digital good life. The conclusion will also draw out the common normative ideas found in these *specific insights* in order to articulate *general insights*. These two sets of insights will then serve to answer the main research

question: **What normative insights can a Confucian philosophical approach contribute to the pursuit of the digital good life?** Besides answering the main research question, the conclusion shall also identify other insights generated by the research, including areas of future inquiry suggested by the research's limitations. The conclusion will also articulate the societal relevance of the dissertation and provide a final reflection.

Chapter 5. Conceptualizing Conceptual Resilience for Living Well with AI: Lessons from Confucian Philosophy[92]

5.1. Introduction

AI systems[93] have become a prominent feature of human existence, both in terms of how they have permeated most parts of everyday human life and the global reach of their influence. This has made them a salient example of "socially disruptive technologies," technologies that can challenge and change the normative practices and relations constitutive of the social sphere (Van de Poel et al., 2023). For instance, as Pak-Hang Wong points out (2021), online job platforms provide the same job recommendations to members with similar profiles due to their AI-powered algorithmic processes. These recommendations, in turn, are based on members' search history. For Wong, this makes an individual member co-responsible for the job recommendations that other members receive. Such additional responsibility, however, would be absent in a situation lacking the AI-driven job platform. Thus, according to Wong, by introducing a novel situation, the AI-driven job platform disrupts the conventional way responsibility operates.

As with other *socially* disruptive technologies, AI systems also have the potential for *conceptual* disruption. This refers to phenomena that challenge (the usage of) concepts in a way that renders concepts inadequate in fulfilling their function (Marchiori &

---

92 This chapter is based on a paper co-authored with Samuela Marchiori, with the author of this dissertation as the first author. The paper has not yet been submitted to a journal. The co-author has given her permission to use this paper for the dissertation.

93 We follow the European Commission High-Level Expert Group on Artificial Intelligence which defines AI systems as: "Software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions" (AI HLEG 2019, p. 36).

Sharp, 2024).[94] Conceptual disruption is not necessarily bad or harmful. However, for example, AI systems' capacity for conceptual disruption might affect important concepts involved in AI ethics and governance. This is concerning because the disruption of these concepts can hinder the formulation of ethical or legal guidelines that can protect people's rights and well-being in relation to AI systems. Take the example of "privacy." AI-driven technologies, such as the aforesaid online job platform, base their decision-making on data from *ad hoc* groups (Mittelstadt, 2017; Wong, 2021). These groups are composed of individuals that AI systems have brought together under a particular classification. The data drawn from these groups can include confidential information about the individual members. However, since AI systems draw data from the group instead of the individuals, the conventional understanding of privacy in the ethical and legal fields – which conceives privacy as attaching to individuals – is ineffective in protecting these individual members. This issue can be understood as an instance of conceptual disruption originating from AI systems. In this particular situation, the concept of privacy is disrupted because it no longer normally fulfils its function of representing the idea of safeguarding personal information.

In response, scholars such as Brent Mittelstadt propose "group privacy" as a novel concept that can be used to address the problem (2017). Mittelstadt's solution can be considered a "curative" approach to conceptual disruption in that it involves an intervention *after* the disruption has happened. In the literature of conceptual engineering, this approach appears to be the dominant one.[95] This paper introduces an alternative strategy to address AI-driven conceptual disruption. We propose engineering concepts that are resistant to conceptual disruption. This would be a "preventive" approach since we seek to preclude the occurrence of conceptual disruption in the first place.

To capture a concept's ability to resist disruption, we coin the term "conceptual resilience." This paper provides a preliminary account of conceptual resilience and

---

94 In this paper, we draw from Marchiori and Scharp's (2024) account of conceptual disruption. On this account, conceptual disruption affects concepts both individually and collectively (disruption of conceptual clusters and conceptual schemes), and manifests differently depending on the level at which the disruption occurs. For the purpose of this paper, we will focus our attention on conceptual disruption affecting individual concepts.

95 Even without explicit reference to conceptual disruption — which, it should be emphasised, is a newly introduced concept — conceptual engineering projects seem to focus predominantly on curative interventions (e.g., "Woman" in Haslanger, 2000, 2012; "Truth" in Scharp, 2013; "Misogyny" in Manne, 2017).

ways to formulate resilient concepts. Our account focuses on two main facets of resilience in the literature: stability and adaptation. From these two facets, we articulate two distinct modes of conceptual resilience, which, in turn, can serve as models or heuristics for designing resilient concepts. These two modes are conceptual-resilience-as-immutability (CRI) and conceptual-resilience-as-adaptation (CRA). We provide examples for each to give a more detailed idea of what designing for CRI and CRA might look like. For CRI, we argue that Björn Lundgren's (2024) recent attempt to design a stable concept can be an example. For CRA, we formulate our own example. Drawing inspiration from Confucian philosophy, we provide a rough and preliminary sketch of what we call an "indicative concept," which we believe exhibits the qualities of CRA. We hope this sketch can serve as inspiration or a basis for future conceptual engineers considering designing concepts with CRA in mind.

This paper thus contributes in four ways to the literature on conceptual disruption. First, it puts forward the idea of conceptual resilience as a means of addressing conceptual disruption. Second, it articulates CRI and CRA as modes of conceptual resilience. Third, it provides a concrete example of what a concept exhibiting CRA might look like. Finally, it introduces resources from a non-Western philosophical tradition to the discussion of conceptual disruption by using Confucian ideas.

Our paper proceeds as follows: In section 5.2, we motivate our idea of conceptual resilience by situating it in the current discourse on conceptual disruption. Section 5.3 provides an account of conceptual resilience and its two modes of CRI and CRA. Section 5.4 provides concrete examples of how CRI and CRA might look by citing Lundgren's work and our example of an indicative concept, respectively. Section 5.5 shows how concepts exhibiting CRI and CRA can be helpful in issues involving AI governance and ethics. In particular, we show how (re)designing the concept of "friend" to be an indicative concept might contribute to the well-being of users when it comes to the issue of whether AI can be friends with humans – an issue that arguably disrupts the concept of friend or friendship. Section 5.6 contains our conclusions.

5.2. From Conceptual Disruption to Conceptual Resilience

Recent work on conceptual engineering in the philosophy of technology has increasingly turned its attention to the phenomenon of conceptual disruption, understood as the interruption in the normal functioning of concepts (Marchiori & Scharp, 2024). A concept's function is "the pattern of using the concept for certain purposes in certain situations" (Marchiori & Scharp, 2024, p.18). For instance, a

concept's function can be to define its referent. According to the notion of conceptual disruption, a concept is disrupted when it is unable to fulfill its functions adequately. The starting point of such discussions is the assumption that technology has an impact on the socio-technical landscape in which it is designed, developed, and deployed. Introducing new technology may put pressure on the conceptual repertoire involved in that landscape. Concepts can become obsolete (e.g., losing relevance), contested (e.g., failing to serve as a shared epistemic or normative tool), or misleading (e.g., introducing distortions into discourse). This can lead to, among other things, breakdowns in communication, epistemic confusion, and normative uncertainty. This can be seen in the abovementioned example of AI systems that challenge the concept of privacy and its normative implications, such that a new concept of "group privacy" might be required.

Focusing on conceptual failures and breakdowns has proven helpful in diagnosing when concepts become obsolete, inadequate, or problematic, prompting interventions to restore conceptual adequacy. Indeed, much of the existing literature on conceptual disruption has focused on identifying cases of disruption and proposing strategies for restoring conceptual adequacy through conceptual engineering interventions, whether through refinement, replacement, or abandonment.

However, to the extent that the literature on conceptual disruption has so far largely focused on investigating when and how concepts break under pressure, it has failed to sufficiently account for how concepts can remain functionally adequate under pressure. Here, we wish to explore the flip side of the conceptual disruption discussion. That is, if concepts are indeed undergoing pressures — whether technology-induced or otherwise — there may be a benefit in articulating how concepts can resist such disturbances. Again, this can be seen from the example of AI's challenge to the concept of privacy, which has ethical and legal implications. Another example is the imminent advent of artificial wombs separate from the body, which might disrupt the concept of "body" (Hopster et al., 2023). If such disruption occurs and persists, it could cause classificatory uncertainty that might hinder the articulation of ethical or legal guidelines meant to protect the rights and well-being of the users of such technologies. For example, as Hopster et al. state (2023), the question as to whether the shutting off of an artificial womb entails a violation of bodily integrity might not be satisfactorily answered due to conceptual disruption. Thus, investigating how the concept of body can be equipped to deal with disturbances and resist conceptual disruption (i.e., be conceptually resilient) should be of interest, not just to conceptual engineers, but also ethicists and legal experts.

Recent work by Lundgren (2024) shifts the attention in this direction, arguing that some concepts can be deliberately designed to resist conceptual disruption. That is, if conceptual engineers could construct concepts that resist conceptual disruptions, the need for future revisions of these concepts might be mitigated or even eliminated. He calls concepts that can resist conceptual disruption "stable" concepts.[96] Lundgren proposes that such concepts can be constructed by formulating them, or their definitions, at a sufficiently high level of abstraction. This, in turn, would make the concept immune to counterfactuals that could make the definition inadequate and thereby cause conceptual disruption.[97]

Lundgren's model provides a compelling account of how some concepts can be (re-) engineered to resist disruption. However, we believe it does not fully capture the plurality of conceivable ways concepts can be engineered to prevent disruption. We thus build on Lundgren's ideas in two ways. First, we introduce a concept that attempts to represent the multiple ways concepts can resist conceptual disruption, of which Lundgren's model is just one. We call this concept "conceptual resilience." Second, we introduce two modes by which conceptual resilience manifests. The first is conceptual-resilience-as-immutability (CRI), of which Lundgren's model is an example. The second is conceptual-resilience-as-adaptation (CRA). In the following section, we shall define conceptual resilience and explain why we consider CRI and CRA to be two of its modes.

5.3. Conceptual Resilience and Its Modes of Immutability and Adaptation

We define conceptual resilience as the ability of concepts to resist conceptual disruption by maintaining continuous functional adequacy despite tensions, pressures, or other disturbances. Conceivably, conceptual resilience can manifest in a variety of different possible ways. This paper, however, focuses only on two modes of conceptual resilience, namely, CRI and CRA. These, in turn, are based on two prominent ways

---

[96] Lundgren also calls stable concepts "undisruptable" concepts. He introduces a distinction between "stable" concepts — which describes concepts which remain adequate despite disruptions — and "undisruptable" concepts which describes concepts which cannot be disrupted (2024, p.1, n1). However, Lundgren claims to be using both "stable" and "undisruptable" synonymously in his work anyway, and what he means by both is a concept that "can avoid disruption" (2024, p.7).

[97] This will be explained in more detail in section 5.4.1.

resilience has been understood in the literature of different disciplines. These ways are resilience as stability, and adaptation.

Before looking at the aforesaid literature, we shall first define stability and adaptation. The Merriam-Webster Dictionary defines stability as "the quality, state, or degree of being stable" (2025). "Stable" in turn is defined by the same dictionary as "not changing or fluctuating" and as "permanent and enduring" (*Merriam-Webster*, 2025). From these definitions, we can see that "stability" entails negating change, or the ability to remain the same, i.e., "immutability." Conversely, adaptation can be defined as "adjustment to environmental conditions" (*Merriam-Webster*, 2025). As "adjustment" implies, this entails changing to cope with, or fit, the circumstance or situation.

"Resilience" has been described as stability and adaptation in the literature of different fields. For example, psychological resilience has been variously described as the ability of adults to "maintain relatively stable, healthy levels of psychological and physical functioning" (Bonnano, 2004, p. 20) and to "remain psychologically healthy or stable despite the fact that they have been exposed to an adverse event" (de Terte et al., 2014, p.416). However, it has also been described as "the ability to recover from perceived adverse or changing situations, through a dynamic process of adaptation" (Caldeira & Timmins, 2016, p.194) as well as "the ability of an individual to successfully overcome negative conditions and adapt to them even when faced with difficult conditions" (Cuhadar et al., 2014, p. 113). In the disaster-response field, community resilience has been defined in terms of stability and adaptation. Patel et al. (2017) have grouped definitions of community resilience into three. One of these emphasizes resilience as "an ongoing process of change and adaptation" in light of challenges or disasters, while another frames resilience as "an ability to maintain stable functioning," again in the face of disasters (Patel et al., 2017). We believe that these examples in the literature lend support to our intuition of resilience as stability and adaptation.

Ecologist C.S. Holling's (1996) well-known distinction between "engineering resilience" and "ecological resilience" further supports our intuition because these two forms of resilience imply stability and adaptation, respectively. Engineering resilience refers to the ability of a system to deflect disturbance or the time it takes to recover from disturbance and return to its original state of equilibrium. An example of a thing that exhibits engineering resilience would be a bent piece of wood that remains bent (or returns to being bent) even if one were to grip it with both hands and try with all of one's might to straighten it. Based on the definition of stability mentioned above,

engineering resilience implies stability since it is understood as a system's ability to remain unchanged despite disturbances. On the other hand, ecological resilience focuses on the range of possible states where a system can still function while absorbing disturbance (Gunderson, 2000; Cretney, 2014). In other words, ecological resilience is the ability of a system to cope with disturbances by changing states while retaining its essential structure and functions. An example would be a piece of clay repeatedly molded by someone's hands, changing shape multiple times, yet ultimately remaining as clay. Based on the definition of adaptation mentioned above, ecological resilience implies adaptation since it is framed as a system's capacity to adjust to a specific situation. In conclusion, resilience is how a system or thing survives disturbances and retains its original structure and function. However, there are several ways resilience can be manifested. Two such ways have to do with the relation of resilience to change. "Stability" is a type of resilience manifested by *immunity* from change, while "adaptation" is a type of resilience exhibited by the *capacity* to or for change.

These two ways of manifesting resilience can be applied to concepts and how they might resist disruption. As mentioned above, Lundgren's strategy of making a stable concept can be considered an example of conceptual resilience in terms of stability. This will be explained in section 5.4.1. However, we will not use the term "conceptual-resilience-as-stability" because we want to avoid the impression that Lundgren's notion of a stable concept is the *only* way concepts can be resilient in terms of stability. Instead, Lundgren's idea is intended to be only one example of how concepts can be resilient in terms of being stable or unchanging. However, using the same word as Lundgren's, i.e., "stable/stability," might give the aforesaid impression. Thus, we will use the term "conceptual-resilience-as-immutability" or CRI. As for conceptual resilience as adaptation (CRA), we shall, as mentioned, provide our own example, namely, that of the Confucian-inspired indicative concept. This shall be explained in section 5.4.2.

To summarize, conceptual resilience refers to the ability of concepts to resist conceptual disruption by adequately fulfilling their functions despite disturbances. This ability can manifest in various ways. Inspired by what the interdisciplinary literature says about resilience, we focus on two of these ways or modes: conceptual-resilience-as-immutability (CRI) and conceptual-resilience-as-adaptation (CRA). These two modes are not meant to be exhaustive. The following section provides specific examples of CRI and CRA. These examples are also not meant to be exhaustive. The figure below illustrates the relationship between conceptual resilience, CRI, CRA, and the examples of CRI and CRA we provide in the next section.

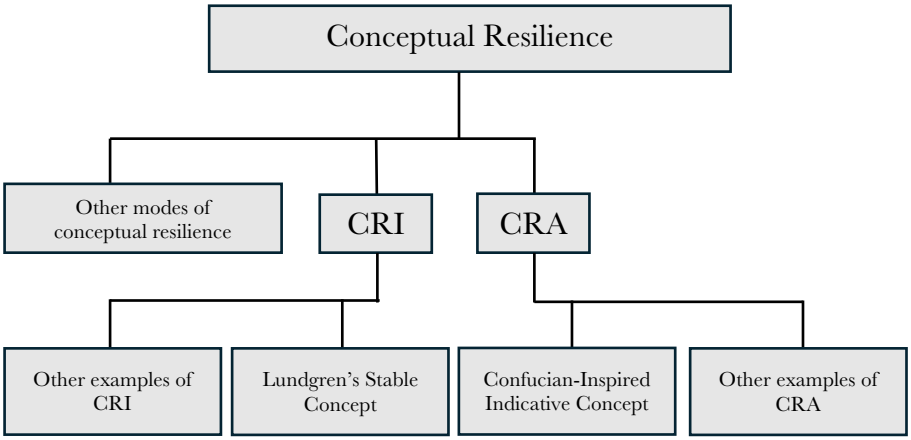


Figure 3

5.4. Examples of CRI and CRA

5.4.1. A Possible Example of CRI – Lundgren's Stable Concept

An example of designing a concept in terms of CRI can be drawn from Lundgren's (2024) attempt to make the concept of "information security" stable or resistant to disruption. Whether Lundgren succeeds in this attempt is not important for our purpose. What is significant is that, for Lundgren, a concept's disruption means a concept's failure to fulfill its function of adequately defining its referent. He does not explicitly say this in his paper. However, this can be inferred from his strategy of making a concept immune to disruption by making the concept's definition "stable against potential counterexamples and challenges" (Lundgren, 2024, p. 5). In other words, a concept can resist conceptual disruption if its definition always adequately captures its referent and never needs to change. Consequently, Lundgren's way of designing a concept against conceptual disruption is to make its definition abstract enough so that no counterexamples can require it to change.

To illustrate this concretely, Lundgren's definition of information security has as its essential part the notion of "appropriate access." That is, information security is achieved if and only if all parties related to a particular piece of information have appropriate access to that information relative to a given stakeholder's needs.[98] Thus,

[98] This is a simplified version of his definition. A more elaborate one is as follows: "An information system S is secure for stakeholder H if, and only if: For every agent A, and every part P of S, A has

information security is defined as the state of affairs where all parties have this appropriate access. Lundgren, however, suggests a counterexample: What if, in the future, technological advancement is such that, even if all parties have appropriate access, a particular kind of technology (like a quantum computer that can override standard encryption) makes the information insecure? (Lundgren 2024, p.6). If that is the case, then the definition of information security as appropriate access fails because the information is still not secure even if appropriate access is obtained.

Lundgren claims that this counterexample, and others like it, are not enough to defeat the definition. This is because appropriate access is an idea at such a high level of abstraction that what it entails can be relativized to the specific contexts where the concept of information security is employed.[99] Put another way, if indeed a technology is developed that can make a piece of information insecure even if the parties involved supposedly have appropriate access, then what is actually the case is that the parties involved *do not have* appropriate access. Appropriate access in this situation would entail something different from what it meant before such a technology was developed, because it would now have to include safety from this technology. In this way, the definition of information security as appropriate access can always adequately capture its referent and, therefore, does not need to change. Consequently, the concept of information security always fulfills its function and is thereby resilient. In this situation, the stability or immutability of the concept (or its definition) is a necessary corollary of the concept's resilience. In short, a concept's resilience is expressed by its immutability. Because of this, we consider Lundgren's example of a stable concept an instance of CRI.

Lundgren's example is also important in that it is, to our knowledge, the only present literature that attempts to explicitly design concepts for resisting disruption. It could thus be said that the only strategy or model so far when it comes to formulating

---

just the appropriate access to P relative to H" (Lundgren & Möller, 2019, p.429, as cited in Lundgren 2024, p. 4). Lundgren intends his definition of information security to be descriptive/ontological, instead of normative (p. 5). That is, it is one thing to ask whether a given information system is secure and another thing to ask if it *should* be secure.

99 As Lundgren states, "The Appropriate Access definition can avoid the challenges of (possible) technologies because it places itself on an abstraction level that is independent of contextual factors" (2024, p. 10). Put simply, Lundgren's strategy is to make the concept, or its definition, more abstract or general. One could remark that this appears to be a rather "commonplace" solution for conceptual disruption in that all concepts can be considered in a very abstract way. However, Lundgren claims that the concept of information security is often defined specifically (and so gets easily disrupted). See: (Lundgren, 2024, p.4). Thus, his suggestion of making the concept more abstract is arguably novel relative to the context in which the concept is employed.

concepts to resist conceptual disruption (or designing concepts for conceptual resilience) is CRI. This raises the question of whether there could be other ways of designing concepts against conceptual disruption. This is not to imply that something is wrong with the CRI model. However, having other models to work with can be helpful since it can provide conceptual engineers with alternatives if the CRI strategy encounters shortcomings or challenges. This leads us to CRA.

5.4.2. A Possible Example of CRA – "Indicative Concept"

Confucius' Treatment of Concepts in the Analects

We will first describe how the *Analects* portrays Confucius' use of ethical concepts. This is because such usage inspires our indicative concept, which we shall detail shortly below. The indicative concept, in turn, exhibits CRA. In the *Analects*, Confucius never defines ethical concepts whenever his disciples ask about them. That is, he never articulates the necessary and sufficient conditions for these concepts. For instance, when asked about the quality of *ren* (仁), "humaneness" or "goodness," Confucius merely describes different aspects of it. For example:

> Zhonggong asked about Goodness [i.e., Humaneness/Ren 仁]. The Master [Confucius] said, 'When in public, comport yourself as if you were receiving an important guest, and in your management of the common people, behave as if you were overseeing a great sacrifice.' (*Analects* 12.2/Confucius, 2003, p. 126).

> Sima Niu asked about Goodness. The Master said, 'The Good person is hesitant to speak.' (*Analects* 12.3/Confucius, 2003, p. 126)

> Fan Chi asked about Goodness. The Master said, 'Care for others.' (*Analects* 12.22/ Confucius, 2003, p. 136)

It could be objected that these descriptions could serve as definitions of *ren*, but there is no indication in the *Analects* that any of the descriptions is meant to be the definition.[100]

---

[100] Another objection is that Confucius' descriptions are operational definitions (Xiao, 2008; Rogacz, 2021). In operational definitions, "the definiendum is defined by a set of operations that results in determining the meaning of the term in question" (Rogacz, 2021, p. 33). For instance, Confucius' response to Fan Chi is that Goodness is defined by fulfilling the operation of caring for others. We grant that some of Confucius' descriptions of ideas can be considered operational definitions. However, this does not take away from our point that classical definitions, understood as singular

Commentators of the *Analects* provide two common reasons for Confucius' refusal to define concepts such as *ren*. The first is that Confucius' aim was never to articulate the definition of concepts to his disciples. Instead, he wanted to describe these concepts in a way that could help them morally progress in the specific situation in which they found themselves. When talking to Sima Niu, for instance, Confucius identifies *ren* with hesitancy in speaking because he believes that the former's loquaciousness is detrimental to his particular moral development. In other words, Confucius used ethical concepts, not to primarily describe or define ethical realities, but to help people attain or experience them (Henderson & Ng, 2014).[101]

The second reason is that attempting to define such concepts would be extremely difficult, if not impossible, since the referents of these concepts are too broad to be encapsulated by a definition. As the Neo-Confucian philosopher, Zhu Xi, says, "the principle of *jen* [i.e., *ren*] cannot easily be conveyed with words. If a fixed definition is given, then violence might be done to the all-encompassing nature of *jen* [i.e., *ren*] itself" (*Zhuzi yulei* 20:21b, as cited in Sato 1986, p. 214). The Neo-Daoist philosopher, Wang Bi, also implies that Confucius uses his descriptions of ethical concepts, not to define, but to "gesture" towards ethical realities (Ashmore, 2004). By gesturing, Wang Bi means that such descriptions indicate phenomena but do not attempt to define them exhaustively. Moreover, these descriptions are situational or context-sensitive in that they highlight features of the referred phenomenon that are salient to the "gesturer" or the "indicator." This explains what Confucius does in the *Analects* since he does not aim to describe ethical realities exhaustively. Moreover, the way he describes these realities is such that he highlights the aspects of these realities that are important in his task of helping his interlocutors grow morally.

The Indicative Concept and Its Features

Confucius' way of using or relaying ethical concepts in the *Analects* can inspire a concept that exhibits CRA. We call this concept an "indicative concept." The basic function of an indicative concept is to refer to its referent. However, the way that the

---

exhaustive formulations of the necessary and sufficient conditions for a thing to obtain, are not in the *Analects*.

101 Compare Steven Geisz's (2008) distinction between the descriptive-representative and pragmatic-regulatory view of language in Classical Chinese thought. The former view of language considers language as mainly a means for adequately describing reality, while the latter view sees language as primarily a means for guiding behavior. According to Geisz, some prominent scholars of Classical Chinese philosophy, such as Chad Hansen, claim that Classical Chinese thinkers adhered to the pragmatic-regulatory view of language.

indicative concept refers to its referent is not through a definition, i.e., a set of propositions that identify the necessary and sufficient conditions for the referent to obtain. Instead, the indicative concept only possesses non-essential and non-exhaustive content about the referent that proves salient in the context in which the concept is employed. This context-sensitivity means that the features of the referent identified by the content need not be the same, or have the same salience, for everyone in all situations.[102] In other words, the indicative concept's role is to merely "point" (hence its name) at contextually salient features of the phenomenon it refers to (or to contextual elements that could actualize the phenomenon), rather than to "encapsulate" the entirety of the phenomenon through a definition.[103]

Another significant feature of the indicative concept is its ethical purpose. Just like how Confucius used concepts in the *Analects* to lead his interlocutors towards moral goodness, the primary purpose of an indicative concept is to assist people in attaining flourishing or the "good life." This explains why the indicative concept is not designed to define its referent exhaustively, and why it is context-sensitive. As mentioned, commentators of the *Analects* interpreted Confucius' disinterest in providing universal definitions of ethical ideas as due to his purpose of helping his interlocutors morally advance in their particular circumstances. This implies that universal definitions are too abstract to fulfil this purpose and that what works more effectively are articulations of these ethical ideas that are specific to people's situations. The indicative concept works under this logic; it does not claim to establish a universal definition of its referent but only includes content that indicates how the referent relates to the good life in a particular context. For instance, if the concept of "friendship" were an indicative one, it would have content that suggests how friends should treat each other in order to live well. However, this content is not meant to be true for all people in all circumstances. Instead, it merely indicates how a friendship can contribute to living well for people in the context where the concept is being used. Even within such a context, the content of an indicative concept should not be understood as stating definitive or absolute claims, since that context could always change.

---

102 The context-sensitivity of the indicative concept does not entail relativism, i.e., the referent of the concept has no objective reality since its features depend only on the context. Instead, the context-sensitivity implies something more akin to moral particularism, i.e., the referent is objectively real. However, its manifestations differ from one context to another.

103 The term "indicative concept," and its contrast to a concept that defines, is inspired by François Jullien's description of Confucius' use of propositions in the *Analects* as a means to indicate realities instead of defining them (2000).

It could be objected that if an indicative concept does not have a definition, it could include anything because no standard or principle determines its content. However, the content of an indicative concept is determined by its ethical purpose. Whatever does not help the "users" of the indicative concept attain flourishing or the good life is not included in the concept's content. This imposes two particular limitations on what can be considered the indicative concept's content. The first involves the conception of flourishing. If the conception of flourishing that the indicative concept depends on precludes certain content, then the indicative concept would likely preclude this content as well. Let us use the concept of friendship again to illustrate this. If a conception of flourishing does not include the state of affairs of "being manipulated," then the indicative concept of friendship also excludes that state of affairs. The second limitation concerns the specific circumstances of people. If the potential content of an indicative concept does not prove helpful – and if it especially proves detrimental — to a person's attainment of the good life in their particular context, then it need not be included. Again, using the example of friendship, such a concept could include the proposition that "friends should often meet each other in person." However, suppose this proposition is unhelpful for a person, perhaps because they are in a situation where they cannot physically meet with one they consider a friend. In that case, the proposition need not be included in the concept of friendship for that person.

Another partial answer to the problem of indicative concepts having no limiting principle for their content is the following: the indicative concept might at the outset already have content about the referent based on the context in which it is situated. When evaluating whether future potential content should be included within the concept, the criteria for such evaluation could be whether the potential content in question has a substantive connection or resonance with the majority of already included content. Moreover, the decision to include potential content should involve dialogue between the stakeholders within the context, particularly between those who suggest the new potential content and those who disagree with them. This is inspired by Wang Bi's implicit idea that one should exercise *shu* (恕), sympathetic understanding or concern, when interpreting descriptions that are meant to be "gestures" (like Confucius' descriptions of ethical concepts above) (Ashmore, 2004). That is, in trying to understand an "indicative description," one should attempt to put oneself in the place of the person who "indicated" and ask why that person came up with that description. Similarly, suppose a party promotes a novel meaning for an indicative concept in a particular context. In that case, people in that context should

try to understand why that party is doing so before judging if such novelty aligns with the concept.

Why an Indicative Concept Exhibits CRA

As mentioned above, an indicative concept does not possess an exhaustive definition of its referent, nor is its function to provide such a definition. Its function is to indicate or provide non-essential statements about its referent. Because of this, an indicative concept can still function in cases where a definitional concept (i.e., a concept whose function is to define) would experience conceptual disruption due to failing to fulfill its function of defining its referent. Again, let us take the concept of friendship as an example. Let us assume that the concept of friendship is definitional and that one of the constitutive parts of its definition – that is, one of the necessary conditions of friendship – is that friends should often meet in person. However, the rise of online communication has challenged this because it has made it possible for people to be friends without meeting in person. Assuming that such online friendships are indeed genuine, then the concept of friendship has become disrupted since it has now failed to provide an adequate definition of its referent that includes all its instantiations. In contrast, if friendship were an indicative concept that does not function to essential or exhaustively define its referent, then a change in the meaning or definition of friendship would not disrupt it.

This ability of the indicative concept to maintain its function despite conceptual challenges or disturbances is why it has conceptual resilience. Moreover, an indicative concept's mode of conceptual resilience is CRA. This can be brought to light if we juxtapose the indicative concept with Lundgren's CRI example. In Lundgren's model, a concept can only be considered resilient if it successfully defines its referent. This success, in turn, means that the concept, or its definition, does not need to change. Thus, Lundgren's stable concept expresses resilience by being immutable or having an immutable definition.

On the other hand, an indicative concept does not have a universal definition that needs to be preserved from change in order for the concept to function. This allows it to absorb its referent's various and changing meanings while still fulfilling its function, thereby exhibiting resilience. Again, this can be shown by the aforesaid example of friendship; if we understand the concept of friendship as an indicative one, even if the meaning or definition of friendship changes, the concept itself will not get disrupted. It will simply include the new meaning within itself. Indeed, the indicative concept can arguably accept both the old and new meanings simultaneously due to the

concept's context-sensitivity. Perhaps in a particular context, the old meaning would be accepted by people since it would be conducive to people's attainment of the good life. In that case, the old meaning is retained in that context. However, perhaps in another context, the new meaning would be accepted since it is the one conducive to flourishing.[104] In that situation, the new meaning would be included in the concept. In all this, the concept would still fulfill its function of indicating its referent in a non-essential and context-sensitive way.[105] Thus, the ability of the indicative concept to avoid conceptual disruption is due to its capacity to incorporate the changes to the meaning or definition of its referent. In short, the indicative concept's resilience manifests as a capacity to change (or accept change), which is then in line with CRA.

5.5. Using Concepts Exhibiting CRI and CRA in AI-Related Issues

To show how CRI and CRA can help in AI-related issues, we will present cases where concepts exhibiting each type of resilience might prove beneficial. We will start with CRI and then move to CRA. However, we will particularly focus on CRA since the discussion above of Lundgren's stable concept already indicates a context where CRI can be helpful. This context is the legal field, which we shall discuss in the immediately following section.

5.5.1. The Use of CRI-Exhibiting Concepts

CRI-exhibiting concepts appear useful for frameworks or practices that require a generally invariable conceptual standard or foundation. That is, the immutability of the concept ensures that it continues to meet its function, which is to provide a consistent foundation for thought, communication, or action in the specific context where the concept operates. As with the case of Lundgren's information security concept, one way the resilience of a CRI-exhibiting concept can manifest is by

---

104 This would not result in contradictory descriptions of the referent because, as mentioned in footnote 102, the indicative concept assumes an understanding of the referent that is akin to the idea of morality in moral particularism, i.e., the referent is objectively real; however, its manifestations differ from one context to another.

105 Engineering the concept of friendship to be an indicative one can thus be useful in these times, when advancing technologies have made it possible for people to have novel *and beneficial* relationships that challenge the conventional meaning of friendship, such as the aforesaid online friendships, as well as "friendships" with AI – a point that we will return to in section 5.5.2.

establishing a generally unchanging meaning of its referent so that it can fulfill its role in consistently guiding reasoning and practice.

Several practices in human society require generally invariable conceptual standards, but perhaps the most salient example is the law.[106] Legal concepts often need considerable consistency because of the law's crucial role in safeguarding citizens' rights and grounding a stable, working, and equitable society, including the international community. This is reflected by the "legal certainty" principle employed in national and international law, such as the EU laws. This principle states that those subject to the law should not be uncertain about their rights and obligations (Tillotson & Foster, 2003). This certainty requires that laws and legal concepts be defined with sufficient clarity and precision so that there will be stability to what they mean, which, in turn, can allow subjects of the law to understand them adequately and consistently predict their implications. In the context of AI systems, the EU's AI Act applies this idea of legal certainty to its AI-related concepts. For instance, the AI Act stipulates that the concept of an AI system "should be clearly defined […] to ensure legal certainty (2024, pp. 4). Thus, the AI Act provides a thorough definition of "AI system" to provide a stable, common understanding of the concept among its many stakeholders.[107]

It could be argued that concepts meant to facilitate legal certainty, such as the concept of AI systems, might better fulfil their purpose if they are engineered to exhibit CRI. This is because CRI-exhibiting concepts, by virtue of their stability (such as the stability of their definitions, as with Lundgren's example), can establish a fixed, enduring meaning of phenomena or ideas that help people avoid the abovementioned uncertainty regarding their rights, obligations, and general standing with the law. Indeed, Lundgren's example of a stable concept is connected to the legal field. The concept he engineers, namely, information security defined as appropriate access, is meant to be a more stable (i.e., its definition is not vulnerable to counterexamples)

---

106 Lundgren's example of a stable concept is also connected to the legal field. The concept that he attempts to make stable, namely, information security defined as appropriate access, is meant to be an improved version of a common definition of information security in the legal arena (Lundgren & Möller, 2019; Lundgren, 2024).

107 The definition is as follows: "'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments" (European Parliament & Council of the European Union, 2024, Article 3).

alternative to a common concept of information security used in US federal law (Lundgren, 2024).[108]

5.5.2. The Use of CRA-Exhibiting Concepts

CRA emphasizes a concept's ability to evolve and adapt in response to changing circumstances. Under this model, a concept is resilient if it can incorporate new information, adjust its boundaries, or modify its application to maintain its relevance and effectiveness. CRA is particularly suited to contexts where flexibility and adaptability are key to maintaining functional adequacy. Concepts operating within rapidly changing environments, such as technology ethics, may require frequent adjustments to remain effective.

A particular issue in the ethics of technology where CRA-exhibiting concepts might be well-suited is the ongoing debate on whether AI tools can be considered friends. We already mentioned online friendships earlier; we now want to focus on the even more contentious discussion of friendships between humans and AI tools. There is voluminous literature on this topic, and it is not the purpose of this paper to thoroughly examine this. Suffice it to say that, in the philosophical literature, the dominant framework for approaching this issue is Aristotelian philosophy. That is, in attempting to evaluate whether AI can be friends with human beings, philosophers often use Aristotle's definition of a friend or definitions inspired by Aristotle's definition.[109] To sum up, Aristotle's definition of friends is that they are "people who must be mutually recognized as bearing goodwill and wishing well to each other"' (*Nicomachean Ethics* VIII.2: 1156a, 4–5, as cited in Kaliarnta, 2016, p. 66). Other significant features of Aristotelian friendship are corollaries of this definition.[110] However, this definition suffices because it already indicates how the Aristotelian understanding of friendship will judge the possibility of friendship with AI. The

---

[108] Lundgren calls this common concept "CIA" to indicate that its core components are confidentiality, integrity, and availability (2024, p.3). Lundgren argues that this concept, or definition, is susceptible to counterexamples because of its rigidity. Thus, he formulates his more abstract concept of information security as appropriate access, as a less vulnerable and more stable alternative.

[109] Examples are: de Graaf (2016), Nyholm (2020), Constantinescu et al., (2022), Bosch et al., (2022), Ryland (2021), Danaher (2019), Fröding & Peterson (2020), and Elder (2017).

[110] For instance, Julia Annas identifies five features that Aristotle considers friends to have. She says: "A friend, then, is one who (1) wishes and does good (or apparently good) things to a friend, for the friend's sake, (2) wishes the friend to exist and live, for his own sake, (3) spends time with his friend, (4) makes the same choices as his friend and (5) finds the same things pleasant and painful as his friend" (Annas, 1989, p. 1).

Aristotelian definition of friendship requires friends to have interiority, or interior faculties like reason and will, by which they can fulfil the actions essential to friendship such as mutually wishing each other well (Bosch et al., 2022; Danaher, 2019). However, AI does not have this interiority. For this reason, or reasons connected to this, scholars drawing from the Aristotelian understanding of friendship have mainly expressed either rejection of or skepticism toward the possibility of AI friendship. (e.g., Bosch et al., 2022; Nyholm, 2020; de Graaf, 2016; Constantinescu et al., 2022; Fröding & Peterson, 2020).

There are, however, scholars who have pushed back against this Aristotelian preclusion of AI friendship by either applying a more loose or limited Aristotelian understanding of friendship (e.g., Ryland, 2021; Danaher, 2019; Kempt, 2022) or by abandoning an Aristotelian approach altogether (Munn & Weijer, 2023). One common and, we believe, praiseworthy reason these scholars have for doing this is their conviction that friendship with AI could be beneficial for people who find themselves in unique circumstances that prevent them from befriending other humans normally. For instance, Helen Ryland, writing during the COVID pandemic, cites the loneliness and isolation that people, especially the elderly, can experience, suggesting that these feelings could be ameliorated or remedied through robot friendship (2021). Although the pandemic has passed, vulnerable groups might still feel a similar isolation due to their particular conditions. Nick Munn and Dan Weijers also cite people who experience past trauma, geographical isolation, and physical disfigurement as those who might have good reasons to befriend AI because of how they are prevented from interacting with other human beings (2023). One could object that online friendships with other people could remedy these difficulties, so there is no need for AI friendship. However, establishing online friendships, particularly substantial ones, might require people to encounter each other visually and/or aurally (e.g., video calls). This, in turn, might be difficult or painful for some people, particularly those who suffer from trauma or disfigurement (Munn & Weijers, 2023). On the other hand, the causes of such difficulty or pain might not be present at all or to a lesser degree when interacting with AI.

Another reason for challenging the "Aristotelian monopoly" on the issue of AI friendship is that there are alternative ways of understanding friendship from other philosophical traditions. These alternatives might be just as effective in conceptualizing friendship as the Aristotelian tradition, and could resonate more with some people. For example, there are Confucian and Buddhist views on friendship (Lu, 2010; Khetjoi, 2019), which might be more relatable for people who personally

adhere to or are inspired by Confucian or Buddhist ideas or for those who live in societies where Confucianism and Buddhism have major cultural prominence. Furthermore, just as European philosophical traditions inspire people from all cultures, non-Western philosophical traditions can do the same, and have done so in the past. These alternative perspectives on friendship might have a response to the question of AI friendship different from the Aristotelian one. For instance, from a Confucian view, personhood is contingent, not on some interior faculty humans possess, but on whether a being ostensibly engages in moral cultivation and helps others to do so (Zhu, 2023). From this, a case could be made that to be a friend, from a Confucian perspective, does not require having the interior faculties humans typically have. Instead, one only needs to ostensibly express attitudes and behaviors fitting of a friend. This might lead to a greater likelihood of accepting AI as friends than the Aristotelian approach.

From the discussion above, it should be clear that the question of AI friendship challenges the concept of "friend," in that competing definitions of friend are simultaneously being promoted. What can be done? We could go the way of Lundgren (i.e., the CRI route) and propose a definition of friend broad enough to encompass all the competing definitions. However, there is a danger that this immutable definition might be too abstract and unhelpful. Moreover, it remains a question whether such a definition, no matter how abstract, can be formulated, considering that the various understandings of friend or friendship might be irreconcilable because of their different philosophical presuppositions.

Instead of following the above strategy, we believe proposing an indicative concept (i.e., the CRA route) might be more beneficial. Engineering the concept of friend into an indicative one would allow the parties above to have their understanding of the concept of friend legitimized without worrying that they have to compete with other definitions. This is because, again, an indicative concept's function is not to formulate a single, exhaustive definition. Thus, it could incorporate these competing definitions of friend because these definitions would be considered merely as non-exhaustive descriptions of friend that appear salient in the specific context in which these definitions were formulated. Because of the continuing advancement of AI, the meaning of friend would likely face future challenges, including possible new descriptions or articulations of it. Again, the indicative nature of the concept of friend would allow it to include these future descriptions. Of course, as mentioned above, the indicative concept cannot accept just any content. Instead, its content should be conducive to the flourishing of its users and should have some substantial relation to

the concept's other content. This, however, would still allow people who genuinely experience AI companionship as part of living well to consider AI as friends.

Indeed, it is possible (whether right now or in the near future) that some people sincerely consider companionship or friendship with AI as a component of their flourishing. Without denying the possibility that these people could be mistaken, it nevertheless seems hasty for philosophers to dismiss their claim of friendship with AI because it does not square with the conceptual framework they are using (which, if it is Aristotelian, can also be considered somewhat dated). These people, who already feel isolated due to their situation, might experience a second layer of isolation – even marginalization – in that they find little or no sympathy from the philosophers or experts who claim to know how AI and humans like them should relate. Something similar applies to those who consider AI as friends based on different philosophical (or cultural) ideas from those of the dominant Aristotelian approach. They might also feel isolated or alienated if the "philosophical consensus" that AI cannot be a friend is based on a concept of a friend drawn from only a particular philosophy, one that they are not familiar with or might not agree with.

Thus, making friend an indicative concept might benefit not just philosophers or ethicists but laypeople as well, since such a concept would be adaptable enough to accommodate and do justice to their distinct understanding of the concept of friend, thereby also giving them a voice in this issue of AI friendship. Moreover, having an indicative concept that honors both the understanding of experts and laypeople could be conducive to promoting solidarity between them since it can give them the impression that their various ways of understanding the concept, though different, might indicate similar or the same things. In other words, although their present understandings might be at odds, they are nevertheless engaged in the same endeavor of trying to comprehend and articulate a common concept or idea. However, suppose there are still strong disagreements about the concept of friendship between these two parties despite using an indicative concept. In that case, dialogue between them should be enacted, in line with Wang Bi's aforesaid idea of *shu* or sympathetic understanding. Considering all this, then, using an indicative concept (and sympathetic understanding) in situations such as these might not only prevent conceptual disruption but also possible – and unnecessary – social friction.

5.6. Conclusion

In this paper, we suggest a way to address conceptual disruption in the context of developing AI technologies by introducing the idea of "conceptual resilience." Conceptual resilience refers to a concept's ability to resist or be immune to conceptual disruption. The idea of conceptual resilience helps fill a number of gaps related to the conceptual disruption literature. One gap is that, whereas attempts to address conceptual disruption have been chiefly "curative" in that they are meant to be enacted after disruption has occurred (e.g., creating a new concept), conceptual resilience points to a "preventive" approach where concepts are engineered so that they are immune to disruption in the first place. The only existing literature that attempts a preventive approach, i.e., Lundgren's (2024), does so by attempting to engineer a stable concept. However, Lundgren's strategy for engineering a concept against conceptual disruption is only one possible method. It is conceivable that there are other ways of designing concepts to resist conceptual disruption. This, then, is another gap, which we fill with the idea of conceptual resilience. That is, conceptual resilience is an idea that is meant to represent the plurality of ways concepts can be engineered to preclude disruption.

We define conceptual resilience as the ability of concepts to resist conceptual disruption by maintaining continuous functional adequacy despite tensions, pressures, or other disturbances. To clarify our idea of conceptual resilience, we show how it could be understood in two modes based on the interdisciplinary literature on resilience. Resilience in the literature has been understood in terms of stability and adaptation. Based on these two understandings, we articulate the two modes of conceptual resilience: conceptual resilience-as-immutability or CRI, and conceptual resilience-as-adaptation or CRA. However, to maintain the idea that conceptual resilience represents the totality of ways concepts can resist disruption, we qualify that CRI and CRA do not exhaust the possible modes of conceptual resilience.

CRI and CRA can serve as models or heuristics for future conceptual engineers in their attempts to formulate resilient concepts. In line with this, we give two specific examples to more concretely show how concepts can be designed to exhibit CRI and CRA. For CRI, we pick Lundgren's attempt to create a stable concept of "information security." In Lundgren's case, making a concept resilient means ensuring that its definition is unchanging or immutable. By having its resilience necessarily linked to its (or its definition's) immutability, Lundgren's stable concept thereby expresses CRI. For CRA, we give our own example of an indicative concept inspired by Confucian philosophy. An indicative concept exhibits CRA because it avoids disruption by being

flexible enough to admit multiple definitions or meanings brought about by conceptual challenges and disturbances.

Finally, we indicate where concepts that exhibit CRI and CRA could be helpful in AI-related issues. CRI-exhibiting concepts can be useful when it comes to AI issues that involve the legal context. This is because concepts in legal frameworks, especially those binding internationally, often require considerable consistency or stability to ensure citizens' rights and welfare. On the other hand, CRA-exhibiting concepts, such as our indicative one, might be more useful in contexts where existing concepts, such as "friend," might need more looseness to accommodate the unique and potentially beneficial state of affairs that could arise due to the advancement of AI technology.

We intend for this paper to serve as a point of departure for conceptual engineers who might wish to design concepts that can be resilient against conceptual disruption. That said, the ideas about conceptual resilience we introduce here are only a preliminary and basic sketch. We expect that they will require further finessing and enrichment as future scholars gain more insights into how concepts can resist conceptual disruption. In other words, our ideas are not "foolproof," as there are still questions or issues regarding their conception and operation.

For example, when a discussion between different parties is required to incorporate new content into an indicative concept, one could argue that conceptual disruption has already occurred, which is the very reason for such a discussion. This, in turn, puts into question whether an indicative concept truly possesses conceptual resilience. Related to this is the possible scenario where an indicative concept has been disrupted in one context, while in another it has not. In such an event, has conceptual disruption indeed happened for the concept? Moreover, what does this say about the concept's resilience?

Another possible issue is the indicative concept's similarity to other definition-less concepts, such as Wittgenstein's "family resemblance" and the "prototype" idea of concepts in cognitive science. One feature distinguishing the indicative concept from these other concepts is its ethical purpose. As mentioned above, the indicative concept's primary purpose is to help people achieve the good life in their specific situations. This purpose is not a typical feature of the family resemblance and prototype idea of concepts. Nevertheless, one could still ask if the indicative concept can be considered a mere subset or species of these other kinds of concepts.

Without downplaying the significance of these issues, we nevertheless leave them for future research (whether our own or that of others) to address.

Chapter 6. Conclusion of the Dissertation

This dissertation explored the normative insights that a Confucian-inspired philosophical approach can contribute to the question of the digital good life. The question of the digital good life asks how people can attain or maintain the good life in relation to digital technologies. Drawing inspiration from Confucian philosophy, this dissertation addressed two salient gaps or concerns in the philosophical discourse on the digital good life. The first is the potential individualistic bias inherent in dominant theories of well-being. Due to the dramatic increase in digital technologies, which have significantly enhanced people's connections with themselves and the world, attaining the good life in a digital context has, to some extent, become dependent on these connections. However, the primary focus on the individual that the dominant philosophical theories of well-being have today is in danger of failing to capture such connections or relations. Confucian philosophy offers a promising resource for addressing this issue, owing to its relational conception of human life and flourishing. That is, using a Confucian-inspired philosophical approach can yield ideas or insights that could more effectively help people live well with digital technologies, because such a philosophy can do justice to the importance of relations in the digital good life.

The second gap is the lack of significant representation of non-Western philosophies in the philosophical discourse of the digital good life. This lack is problematic for three reasons. The first is that the lack of non-Western philosophy, or the dominance of Western philosophy, in the discourse of the digital good life, might make the issue of the digital good life more challenging to understand for people who prefer non-Western philosophies or live in cultures influenced by these philosophies. The second is that excluding non-Western philosophies from the discourse on the digital good life might delay or prevent the acquisition of beneficial ideas that could have been gained if they were included. Finally, including non-Western philosophies in the aforementioned discourse is a matter of justice, as contemporary academic philosophy is often dominated by Anglo-European thought, which marginalizes non-Western philosophical traditions. Approaching the digital good life with Confucian philosophy can address the first gap because it is a non-Western tradition whose influence remains salient in some regions of the world, particularly in East Asian nations. Moreover, as mentioned above, and addressing the second gap, Confucian

philosophy is a promising source of helpful insights or ideas for the digital good life, which might not be easily attained by the more individualistic-leaning Western theories of well-being. Lastly, Confucian philosophy can address the third gap simply because it is a non-Western philosophy.

With these considerations in mind, the dissertation thus asked the following main research question: **(RQ) "What normative insights can a Confucian-inspired philosophical approach contribute to the pursuit of the digital good life?"** To answer this, the dissertation employed a framework inspired by the Confucian classic, the *Great Learning*, where the digital good life supervenes not only on the individual human but also on the spheres of relations in which they are situated. Based on this framework, four issues were explored in the dissertation, each of which is connected to a different digital technology. These technologies, in turn, influence a sphere of relation that constitutes the scope of the digital good life. The spheres, along with their associated digital technologies and issues, were discussed as follows: In the sphere of the self, the technology discussed was mindfulness apps, and the issue was "McMindfulness." In the interpersonal sphere, the technology was online multiplayer games, with the issue being toxic behavior. The political sphere was characterized by social media technology, and the issue surrounding it was online political polarization. Finally, the sphere of the world had AI systems as the technology, with the issue being the conceptual disruptions they introduce. Confucian philosophy was then applied to these issues with the aim of advancing the digital good life of the users involved.

The application of a Confucian-inspired philosophical approach to these four issues was articulated in the form of four subquestions. These subquestions, each of which was tackled by chapters 2-5, respectively, were the following:

(**SQ1**) **How can conceptual resources from Confucian philosophy inform the efforts of designing mindfulness apps against 'McMindfulness'?**

(**SQ2**) **What normative insights can Confucian philosophy offer when it comes to mitigating toxic behavior in online multiplayer video games?**

(**SQ3**) **How can ideas inspired by the Confucian cognitive and ethical view of emotions contribute to reducing political polarization on social media?**

(**SQ4**) **How can Confucian views on concepts inform the efforts of engineering resilient concepts against conceptual disruption afforded by AI systems?**

The answers to these subquestions provide the *specific insights* that a Confucian-inspired philosophical approach can offer in the pursuit of a digital good life. They are "specific" since they are involved with only a particular issue, technology, and sphere of relation. On the other hand, the common ideas that can be extrapolated from these *specific insights* are the *general insights* that a Confucian-inspired approach can contribute to advancing the digital good life. The *general insights* would apply to all the abovementioned issues and their corresponding technologies and spheres of relations. The *specific* and *general insights* shall constitute the answer to the main research question.

Having said the above, the next section will proceed to answer the main research question. The section will be divided into two. Section 6.1.1 will discuss the *specific insights*, that is to say, the answers to the subquestions. Section 6.1.2, on the other hand, will identify the *general insights* by drawing out the common ideas from the specific insights.

6.1. Answering the Main Research Question

6.1.1. Specific Insights

SQ1: *How can conceptual resources from Confucian philosophy inform the efforts of designing mindfulness apps against 'McMindfulness'?*

'McMindfulness' is the term applied to contemporary digital mindfulness products, such as mindfulness apps, that are seen to have become instantiations of neoliberal governmentality. According to critics of these apps, by lacking a substantive ethical orientation, promoting an atomized and individualistic perspective of the self, and deflecting critical examination of the social structures in one's environment, McMindfulness products reinforce the structural inequities and onerous lifestyle promoted by neoliberalism. Thus, McMindfulness can be an obstacle to living well or the good life.

Chapter 2 argues that the philosophy of Neo-Confucian philosopher Zhu Xi can be a resource for reforming mindfulness against McMindfulness because it has a conception of mindfulness (*jing* 敬) that is (1) integrated in an ethical framework (i.e., the Great Learning/*Daxue* 大學) that considers one's socio-political context as constitutive of one's flourishing, (2) connected to a practice of careful inquiry of one's environs (i.e., investigation of things/*Gewu* 格物), and (3) motivates socio-ethical engagement.

Drawing on Zhu Xi's ideas, several design recommendations for mindfulness apps were presented in chapter 2 that could help these apps counter McMindfulness and its effects. The first and primary feature is a course in a mindfulness app (or a mindfulness app itself) that teaches 'Zhu Xi-inspired mindfulness' (ZIM). ZIM is a mindfulness practice that can instill in users a sense of ethical responsibility for their social context, encouraging them to examine this context for issues or areas where they can improve the quality of life. Other design recommendations include: (1) A functionality in the app where users can type the reflections gained in their mindfulness practice, (2) different modes of communication in the app that can facilitate group activities thereby promoting a sense of community that can counter the neoliberal individualism McMindfulness promotes, and (3) a feature where an app would notify users of significant ethical issues happening in their local socio-political contexts which, in turn, could contribute to users' awareness and sense of responsibility for such contexts.

SQ2: *What normative insights can Confucian philosophy offer when it comes to mitigating toxic behavior in online multiplayer video games?*

Toxic behavior in online multiplayer games consists of "abusive communications directed towards other players [. . .] and disruptive gameplay that violates the rules and social norms of the game" (Beres et al., 2021, p. 1). Such abusive actions can significantly hinder users' experiences of living well. *League of Legends*, a popular online multiplayer game, is known for its toxic behavior. Certain features of *League* facilitate toxic behavior, such as the ranking system and the snowballing phenomenon.

Chapter 3 argues that from a Confucian perspective, some of *League's* features, besides being conducive to toxic behavior, also prove inimical to Confucian values. For instance, the act of killing players, or the champions of other players, runs against the Confucian appreciation for life and disdain for killing. The ranking system and attention-demanding nature of *League* preclude the exercise of the Confucian virtue of sympathetic concern (*shu* 恕) because these features prevent users from imagining and reflecting on how they would feel if they were in the place of other users. Finally, the combination of different features, such as the ranking system and snowballing, instills in users a compelling desire to win and, consequently, an aversion to novel or alternative playstyles that may not lead to victory. This lack of openness and accommodation to difference contradicts the Confucian value of interpersonal harmony, which entails a welcoming attitude to such difference.

Confucian philosophy can inform design features that reduce toxicity in gaming by promoting core values. One feature could encourage toxic players to re-watch their interactions with those they have harmed, providing ample time and "attentional space" for fostering sympathetic concern. Additionally, players could be required to send at least one positive message to a teammate and an opposing player before and after each match, reflecting the Confucian value of ritual (*li* 禮) to enhance interpersonal harmony. Furthermore, to reflect the Confucian reverence for life, the depiction of champion deaths could be removed, while tracking instances where players save teammates instead of just focusing on their kills. This shift would valorize the act of saving lives, redirecting players' competitive instincts towards more positive outcomes. This redirection is aligned with the Confucian idea of harnessing non-moral inclinations for moral purposes. Although reducing the game's valorization of killing might not directly reduce toxic behavior, it may lessen the aggressive competition that could lead to it.

SQ3: *How can ideas inspired by the Confucian cognitive and ethical view of emotions contribute to reducing political polarization on social media?*

Online political polarization poses challenges for cooperation and social harmony (Ziliotti, 2022). Chapter 4 argues that existing strategies to address online polarization often exhibit a "rationalistic bias," downplaying the role of emotions in the process. While diversifying content is thought to reduce bias, research indicates that exposure to opposing viewpoints can heighten polarized feelings. This aligns with the theory of "affective polarization," which states that people's affective attachments to their ingroup lead to hostility toward the outgroup. Thus, users are not simply rational actors; their affective attachments influence their reactions to diverse content.

The theory of affective polarization counters the rationalistic bias by acknowledging the significance of emotions, but may risk viewing them as obstacles to rational discourse. In contrast, cognitive emotion theories, such as affectual intuitionism, emphasize that emotions are integral to perceiving and understanding moral values, rather than mere irrational hindrances. Moreover, emotions can be harnessed to mitigate polarization by fostering an appreciation of values that are inimical to such polarization.

Chapter 4 argues that Confucian philosophy offers valuable insights in this context, emphasizing cognitive emotions through the concept of the heart-mind (*xin* 心) and their role in ethical flourishing. Confucian methods can cultivate cognitive emotions that promote values countering polarization. For example, Mencius'

method of extending moral emotions can inspire a social media design that prompts users to pause, remember, and reflect on past experiences of these emotions (such as care). This can help them apply such emotions when engaging with outgroup content. Additionally, drawing on Zhu Xi, Dai Zhen, and Mencius, similar prompts can ask users to give reverential attention to Confucian views on human nature, thereby motivating moral and sympathetic interactions with others, including those from the outgroup. Furthermore, employing Dai Zhen's idea of human nature, posts could be categorized based on universal human desires, highlighting shared aspirations among users despite political differences.

Finally, Confucian ideas on moral exemplars can be applied to promote "online exemplars" that help users appreciate attitudes and behaviors that counteract polarization. Social media design features, such as badges and comment ranking, can highlight users with admirable attitudes. Though these users might not be wholly virtuous, the Confucian idea of "imperfect exemplars" suggests that even they could inspire others. Indeed, these relatable figures may be more motivational than fully virtuous individuals. Additionally, the concept of "influence modeling" encourages individuals to strive to be moral exemplars, recognizing their impact on others. Influence modeling can also inspire a prompt that encourages users to reflect on their moral potential, motivating them to act in ways that foster moral goodness in their interactions.

SQ4: *How can Confucian views on concepts inform the efforts of engineering resilient concepts against conceptual disruption afforded by AI systems?*

Chapter 5 discusses how the advancement of AI systems raises the potential for conceptual disruption, which can lead to uncertainty about the meaning or function of key concepts, particularly in the field of AI ethics. This is concerning as it may hinder the development of ethical guidelines vital for protecting people's rights and well-being. One approach to combat this disruption is to engineer concepts that resist it. Lundgren (2024) has proposed creating immutable definitions for concepts, but this is just one method. Chapter 5 introduces the concept of "conceptual resilience," defined as a concept's ability to maintain functional adequacy (i.e., remain undisrupted) amid various pressures. The chapter outlines two modes: the first is conceptual-resilience-as-immutability (CRI), as exemplified by Lundgren, where conceptual resilience is tied to a concept's unchanging core component, such as its definition; the second is conceptual-resilience-as-adaptation (CRA), in which

conceptual resilience is determined by a concept's ability to adapt flexibly to disturbances.

Confucian philosophy contributes to the notion of CRA through the idea of a "indicative concept," where ethical concepts do not provide exhaustive, fixed definitions but instead operate contextually to aid moral growth. Such concepts adapt to disturbances, such as changes in definition or meaning, without losing their relevance or functionality. Nevertheless, novel definitions or meanings ascribed to indicative concepts might not be acceptable to other "users" of these concepts. In this case, these users should engage in dialogue with the promoters of the new meaning, striving to understand their intent for introducing this novelty. This aligns with the Neo-Daoist philosopher Wang Bi's notion of understanding gestural terms through sympathetic concern (*shu* 恕).

The notion of an indicative concept can enrich the AI ethics discourse, for example, concerning the debate on whether AI can be friends with humans. While many scholars reject the idea based on traditional Aristotelian definitions of friendship, others argue that AI can provide companionship to those who are unable to connect with others. This has led some to propose alternative definitions that accommodate AI as friends. Consequently, a fundamental dispute arises between pro and anti-AI friendship advocates, each holding irreconcilable definitions. By framing "friend" as an indicative concept, both sides can validate their perspectives without competing, which may help individuals who rely on AI companionship feel acknowledged and supported, ultimately enhancing their well-being. A broad, indicative understanding of friendship may foster greater rapport between differing groups, revealing commonalities despite differing views. From a Confucian standpoint, this social harmony contributes to overall well-being and the good life.

6.1.2. General Insights

The previous section identified the normative insights that a Confucian-inspired philosophical approach can provide regarding four ethical issues, each related to a particular digital technology that respectively influences one of the four spheres of relations constituting the good life. This section will highlight additional, cross-cutting normative insights that provide further answers to the main research question. That is to say, drawing from the answers to the four subquestions, this section identifies four overarching or general normative insights that a Confucian-inspired philosophical approach can offer when it comes to live well in relation to digital technologies. These

general insights are (1) the crucial role of pause and reflection, (2) the significance of engaging the cognitive, affective, and sensory faculties of users, as well as (3) "meeting users where they are," when it comes to strategies for advancing the digital good life, and (4) the importance of making users aware of the shared nature of the digital technology and establishing cooperative and/or ethical relations with other users.

The following four insights will be discussed consecutively below. The discussion for each of these insights shall be in two parts. The first part will describe how the insight is present or reflected in each of the answers to the subquestions. The second part will elaborate on the insight and discuss its implications for advancing the digital good life. After these discussions, a synthesis of the four insights will be provided. This will yield a final insight, namely, (5) that digital technologies have the potential to support the entire process of what is called 'self-cultivation', through which people can attain the good life.

Insight 1: The Crucial Role of Pause and Reflection

The Insight as Found in the Answers to the Subquestions

The Confucian-inspired approach to the four subquestions, as discussed in the chapters of this dissertation, generally encourages pausing from one's daily activities to reflect on ethical and prosocial considerations. For example, Zhu Xi-inspired mindfulness (chapter 2), or ZIM, encourages mindfulness app users to set aside their current activities so that they can reflect on ethical ideas and the ethical quality of their environment with a tranquil and concentrated mind. While mindfulness apps aim to encourage breaks from daily life, ZIM's distinct purpose is its ethical emphasis. The recommendations for mitigating toxic behavior in online multiplayer games (chapter 3) also imply pause and reflection. The re-watching feature interrupts players' immersion in the game setting and invites them to reflect on their toxic behavior and its effects on others. The feature that requests players to say something positive to fellow players before and after each match similarly serves as a brief pause from playing the game, providing an opportunity for them to reflect on the importance of harmonious interactions.

The Confucian-inspired suggestions for reducing social media polarization (chapter 4) also entail taking a break or pause from social media immersion to reflect on ethical ideas. For instance, the prompts inspired by the method of "reverencing human nature" urge users to pause and reflect on how they should interact with the partisan outgroup by contemplating ethical feelings such as care and the ethical implications of Confucian conceptions of human nature. The Confucian-inspired

indicative concept (chapter 5) also implicitly encourages pausing and reflecting. This is because when the concept is given a meaning that might be difficult to accept, people should try to understand why the party that assigned such a meaning did so. This entails treating the party sympathetically, which requires allotting time to reflect on their intentions.

Elaboration of the Insight and Its Implications for Advancing the Digital Good Life

Pausing and reflecting on ethical considerations has been a salient part of the Confucian tradition. This reflection can come in different forms. It may involve considering one's ethical duty to others (*Analects* 1.4), putting oneself in another's place sympathetically (*shu* 恕), or contemplating one's inherent moral potential (Mencian extension). One reason why pausing and reflecting on ethical considerations is important is that it can be what (re)directs people towards the moral path and motivates them to follow it. This ensures that they remain aligned with the moral path amidst the hustle and bustle of daily affairs, which could otherwise distract them from the path or even lead them astray from it.

In the context of pursuing the digital good life, pausing to reflect on ethical considerations can prove crucial. Digital technologies, with their attention-capturing designs, can easily become an element of one's daily life that distracts one from the ethical demands of the present moment, thereby preventing one from living a good life. What this insight suggests, however, is that digital technologies can also serve the opposite purpose. They can connect people, and if well-designed, facilitate breaks from the churn of routine, allowing users to evaluate their current experience from an ethical perspective and determine how to conduct themselves in line with the good life. This is especially important in the context of the online environment. Due to factors such as anonymity and the absence of physical presence, online communication can encourage users to express abusive behavior towards others, which they would not do in person.[111] Thus, there is a greater need in the online realm for reminding people to pause and reflect on their attitudes and conduct.

[111] This phenomenon is what scholars refer to as the "Online Disinhibition Effect." See: Suler (2004).

Insight 2: The Significance of Engaging the Cognitive, Affective, and Sensory Faculties of Users

The Insight as Found in the Answers to the Subquestions

The Confucian-inspired approach to the four subquestions recommends methods of advancing the digital good life that engage users' cognitive, affective, and sensory faculties. The ZIM program (chapter 2) invites users to contemplate the virtue of humaneness (*ren* 仁), as well as other ethical ideas, concretely. This entails imagining how one might feel when giving and receiving acts of humaneness (*ren* 仁). This act of imagining taps into the users' cognitive ability to understand the virtue. It also taps into their emotions by its attempt to "feel" what the virtue is like. Moreover, the act of imagining simulates sensory experience, thereby tapping into the capacity of the senses or sensory faculties.

The Confucian-inspired approach to mitigating toxicity in online multiplayer games (chapter 3) also taps into users' cognitive, affective, and sensory faculties. The re-watching feature engages a toxic player's sensory faculties by providing them with a visual and aural portrayal of how they mistreated another player. The feature also potentially engages the emotional faculties of the toxic player since the concrete portrayal of their toxicity might evoke certain emotions, such as shame. Furthermore, the feature connects with the cognitive (and possibly emotional) faculties of the toxic player by asking them to put themselves in the place of the bullied player and reflect on how the latter could have felt. In other words, the toxic player is asked to exercise *shu* (恕), or sympathetic concern, towards the offended player.

Using imagination – and consequently the cognitive, affective, and sensory faculties of users – also plays a significant role in the Confucian-inspired recommendations for reducing social media polarization (chapter 4). The various "prompts" inspired by Mencian extension and reverencing human nature invite users to imagine ethical ideas concretely. For example, the Mencian prompt asks users to remember past experiences of care, which simulate sensory experience, thereby tapping into users' sensory faculties. Moreover, these prompts, banking on the notion of cognitive emotions, urge users to "feel" or direct emotions toward these ideas. For instance, the Mencian prompt, which asks users to recall past instances of experiencing care, also invites them to re-experience that care.

Sympathetic concern or understanding is also the reason why people's cognitive, affective, and sensory faculties are potentially involved in the Confucian recommendation for the issue of conceptual resilience in AI governance (chapter 5). Disagreements regarding the meanings assigned to an indicative concept should be

resolved through dialogue, in which different parties attempt to understand each other. This is inspired by Wang Bi's idea that gestural terms should be understood through *shu* (恕)or sympathetic understanding. This, however, suggests that, in trying to understand each other, the aforementioned different parties should imaginatively put themselves in each other's place, since this is what *shu* (恕) entails (Tiwald, 2010). This, in turn, likely involves the use of the cognitive, affective, and sensory faculties, similar to the imaginative exercise mentioned above.

Elaboration of the Insight and Its Implications for Advancing the Digital Good Life

The Confucian tradition's method of cultivating virtue often engages the person's cognitive, emotional, and sensory faculties together. This is due to the holistic understanding Confucian philosophy has of human beings, which does not absolutely divide them into cognitive/rational, emotional, and sensory parts (Seok, 2013). Other methods of cultivating virtue, aside from those already mentioned, illustrate this holistic approach. For instance, Confucians believe that music and ritual performance can significantly influence people's moral dispositions due to their combined impact on the mind, emotions, and senses (e.g., *Xunzi*, ch. 19, 20). Confucians, such as Mencius, also considered the holistic approach more effective in moving people towards moral goodness than methods that engage only one of the faculties, e.g., the cognitive or rational one. This is implied by Mencius' use of his technique of extension – which incorporates cognitive, emotional, and sensory elements through the act of imagining and reflecting – as a likely foil to the method of formal argumentation that the Confucians' philosophical rival, the Mohists, mainly relied on (Slingerland, 2003).

Using interventions that engage the cognitive, emotional, and sensory faculties simultaneously may prove effective in influencing people to conduct themselves in a manner conducive to the digital good life. Indeed, scholarly literature has already implied as much. For instance, one factor that makes it conducive for online users to mistreat each other is that they do not perceive each other's physical presence (Wachs et al., 2019). This diminished sensory exposure and concomitant desensitization to the other person weakens emotions that might otherwise prevent users from mistreating them in person. This supports the importance of engaging users' emotional and sensory faculties when it comes to helping them live ethically in digital space. Similarly, Marin and Roeser (2020) argue that the dominance of textual communication and the limited means of conveying emotions in social media hinder more nuanced emotional expressions. This, in turn, could lead to misunderstanding among users and the prevalence of more easily conveyable, but more polarizing,

emotions. They thus propose design recommendations for social media that can help users communicate their emotions more comprehensively. What Confucian philosophy can add to this is the idea of engaging users' imagination, as the various Confucian-inspired prompts in chapter 4 illustrate. That is, instead of only designing digital features that relay emotional or sensory content to users, features can be made that invite users to use their emotional and sensory faculties by exercising their imagination. This suggests a more active role for users in the pursuit of the digital good life.

Insight 3: Interventions for advancing the digital good life should "meet users where they are" so that they might be more effective

The Insight as Found in the Answers to the Subquestions

A common feature among the Confucian-inspired recommendations for the subquestions is that they "meet users where they are" in order to be effective. "Meeting users where they are" means that these recommendations take into consideration and utilize the user's context, experience, and inclinations. Through this, users can more easily identify or resonate with the ideas or activities that these recommendations invite them to engage in. For example, the ZIM program (chapter 2) aims to help users address the structural injustices of their social environment. However, the initial step for doing this is for users to observe their own lives and those of their nearby relationships. Through this observation, users can become aware of the difficulties that they and the people around them experience. This, in turn, can lead them to identify the structural inequities that may be the source of such difficulties. By making users observe the lives that are already close and (often) important to them, ZIM makes the idea and activity of addressing structural injustice more relatable to users.

The concept of "meeting users where they are" is also reflected in Confucian-inspired recommendations for reducing toxicity in online multiplayer games (chapter 3). Specifically, it can be found in the recommendations for recording the number of times players saved their teammates' lives in the game. This recommendation aims to devalue killing in the game and replace it with a valorization of life. However, instead of merely informing users that killing is bad and valuing life is good, the recommendation takes into account players' desire to showcase their skill. That is, by recording how many times players saved their teammates, the recommendation frames saving lives as a noteworthy achievement, and thereby taps into players' desires for skillfully accomplishing achievements and having them displayed for others to see.

In this way, the recommendation "meets players where they are" since it utilizes the desire for displaying skill that users already have in order for them to appreciate the idea of valuing life.

"Meeting users where they are" is similarly present in the Confucian recommendations for reducing social media polarization, particularly in the Mencian prompt (chapter 4). The purpose of the prompt is to encourage users to show care towards the outgroup when they encounter them or their content on social media. To achieve this, the prompt invites users to remember and reflect on an instance of caring they have experienced. This can then motivate users to show care towards the outgroup. Thus, the prompt taps into an experience that users already have (and likely appreciate) in order to make it easier for them to engage in the act of expressing care. The idea of imperfect exemplars also relies on the idea of "meeting users where they are." Although users who are highlighted as exemplars of respectful online conduct might not be entirely virtuous, their very imperfection might make them more relatable to other users, thereby encouraging the latter to emulate them. Finally, the idea of "meeting users where they are" is also reflected in the notion of indicative concept (chapter 5). One of the aims of the indicative concept is to help its users attain the good life. It fulfills this goal not by providing users with abstract or context-independent content, such as a fixed universal definition of its referent. Instead, it offers content relevant to the user in their specific situation. In this way, the indicative concept aids users in living well by using their context and experience.

Elaboration of the Insight and Its Implications for Advancing the Digital Good Life

For prominent Confucian thinkers, "meeting people where they are" appears to be a common strategy. This is already indicated by the recommendations above, which said thinkers inspire. For instance, Confucius' description of ethical concepts in the *Analects* was tailored to the particular situation of his interlocutors. That is, in teaching others about how to live well, Confucius "met people where they are," because he customized his teaching to their own experience. Another example is the recommendation of utilizing or channeling a person's non-moral inclinations towards moral ends, as suggested in chapter 3 and drawn from Mencius. To reiterate, Mencius tells King Xuan to channel his vicious inclinations, such as the desire for wealth and aggression, into virtuous goals. In doing so, Mencius "met the king where he was" because he took into consideration the king's already existing inclinations instead of simply telling the king to eliminate them. "Meeting people where they are" can be

considered a practical and pedagogical strategy that Confucians employed to encourage people to embark on, and continue walking, the good life.

There are two general cases where the Confucian insight of "meeting users where they are" might prove helpful in pursuing the digital good life. The first is when there are interventions that attempt to promote the digital good life, but in a way that is not relatable to users. An example is the privacy notices that appear whenever users enter a website. These notices supposedly intend to protect users by giving them a choice on how their data will be used. However, how these notices are presented – abruptly appearing on the screen and presenting users with a voluminous amount of text, often about abstract ideas – can prove annoying and may likely discourage users from interacting with them. Therefore, there is a need to make these notices more relatable to users. Perhaps asking users to think of times their privacy has been compromised, and relaying this in a less textual, more sensory manner, might be a more effective approach. The second case where the Confucian idea of "meeting users where they are" might be worth contemplating when digital technologies relate to users but do not ultimately lead them towards the digital good life. An already mentioned example would be mindfulness apps that serve as instantiations of McMindfulness. These apps cater to the need for relaxation and tranquility that people in contemporary society often need. Thus, they can be said to "meet them where they are." However, these apps, if devoid of an ethical and critical framework, might only serve to further embed people in an onerous lifestyle that ultimately hinders them from achieving the digital good life.

Insight 4: The Importance of Making Users Aware of the Shared Nature of Digital Technology and Establishing Cooperative/Ethical Relations with Other Users.

The Insight as Found in the Answers to the Subquestions

The Confucian-inspired approach to the subquestions commonly provides recommendations that help users become aware of the shared nature of their digital technology and encourage them to form ethical and cooperative relationships with other users. The recommended features for mindfulness apps (chapter 2) that enable users to communicate and engage in mindfulness practices with one another are examples of this. These features introduce users to the shared nature of the mindfulness app by connecting them to other users. At the same time, these features introduce users to each other as fellow mindfulness practitioners with whom they can engage in ZIM and pursue the digital good life together. The Confucian-inspired

approach to reducing toxic behavior in online games (chapter 3) also makes users aware of the shared nature of such games, though not in the same way as mindfulness apps. In *League*, players already know that there are other players. However, some of *League*'s features, such as its individualistic ranking system, can prevent players from realizing that their fellow players should be treated with respect. The Confucian-inspired feature that prompts players to interact positively with fellow players before and after matches can address this by fostering attitudes of mutual respect and appreciation between them. That is, the feature encourages players to recognize that other players should be respected by establishing ethical and cooperative relationships among them.

Something similar applies to the Confucian-inspired approach to reducing polarization in social media (chapter 4). Although users of social media are aware of the existence of other users, they can fail to see members of the partisan outgroup as deserving of respectful consideration. Confucian-inspired interventions, such as the Mencian and reverencing human nature prompts, remind users that members of the outgroup are also entitled to sympathy, or at least respect, due to their shared humanity. Put differently, these interventions make users aware that social media is a technology that their ingroup does not, and should not, monopolize. Instead, social media is a space they share with others who might disagree with them yet deserve the same consideration they would want for themselves. This realization enables users to establish more wholesome and ethical relationships with the outgroup. The Confucian ideas on moral exemplars also help users realize the shared nature of social media and establish ethical relations with other users. In particular, the idea of influence modeling – in which users conduct themselves appropriately because they are aware that they can also be an example for other netizens – helps users become cognizant that people like them also inhabit social media and that they have an ethical responsibility towards them.

Finally, the Confucian-inspired indicative concept (chapter 5), although not a digital technology, also helps users become aware of the shared nature of concepts and encourages them to engage in dialogue with others about the content of these concepts. The indicative concept's lack of a fixed, universal definition allows it to accept a greater variety of content from a possibly greater number of people. This can help users realize that the concept is indeed shared with others and that they do not have exclusive possession of its content or meaning. Moreover, the recommendation to engage in dialogue with parties who propose novel but difficult-to-accept meanings

for the indicative concept encourages people to establish cooperative relations for a more harmonious coexistence.

Elaboration of the Insight and Its Implications for Advancing the Digital Good Life

As shown above, making users aware of the shared nature of digital technology can mean two things. The first entails making users aware of the importance of treating others as they would want to be treated, namely, as humans deserving respect and consideration. This idea has its roots in the general Confucian notion of treating others with respect, sympathetic concern (*shu* 恕), and humaneness (*ren* 仁). The second involves literally informing users that other people are also using the technology they are engaging with. This is the case for the Confucian-inspired approach to mindfulness apps. The purpose of informing users that others are using the mindfulness app is to enable them to engage in the ZIM practice and, ultimately, pursue the digital good life together. The idea that the good life should be pursued in collaboration with other like-minded individuals is a prominent feature of the Confucian tradition. Confucius and his disciples were an example of a community that supported one another in their pursuit of living virtuously. Neo-Confucians such as Zhu Xi and He Xinyin also fostered a spirit of collaboration among aspirants of the moral path by encouraging them to discuss ethical ideas with each other to gain a comprehensive understanding of them (Keenan, 2011). Moreover, as mentioned in chapter 2, Neo-Confucians were known to keep diaries documenting their daily efforts to cultivate virtue, which they exchanged with one another to receive helpful feedback.

Given the rise of toxic behavior, extreme polarization, and similar undesirable phenomena in digital space, the idea that users should be made aware that fellow users deserve respectful consideration appears to be an obviously needed recommendation. What is perhaps more interesting is the other idea mentioned above, namely, that users should be made aware of fellow users so that they can pursue the digital good life in collaboration. This idea might come more easily to those using digital technologies designed for multiple users, such as multiplayer games or messaging apps. However, some digital technologies are designed more for solitary use. Nevertheless, these technologies may still have a substantial impact on users' prospects of achieving the digital good life. Therefore, it might be good for users of these "solitary" technologies to form groups through which they could discuss and evaluate whether such technologies are helping them to live well. For example, large language models (LLMs) – which are well-suited for solitary use – are utilized by users to acquire

information. However, the information these models provide is not always correct. Uncritical acceptance of this misinformation can, in turn, affect users' prospects of living well. Thus, it may be beneficial for individual users of the same LLMs to collaborate in assessing the effectiveness of these technologies in providing accurate information, at least for certain areas of knowledge about which these users commonly inquire.

Insight 5: Digital Technologies have the Potential to Support the Entire Process of Self-Cultivation, Through Which People Can Attain the Good Life.

One way to view all four insights is that they indicate different parts of self-cultivation. "Self-cultivation" (*xiu yang* 修养) is a term used in Confucian philosophy to refer to the process wherein one fully develops and exercises one's moral capacities, thereby achieving moral flourishing (Ivanhoe, 2000). Drawing on the *Great Learning*, self-cultivation can be broadly divided into two phases. The first can be considered the "inward" phase in that it involves ethically shaping one's interior faculties (e.g., making intentions sincere, rectifying the heart-mind). The second can be called the "outward" phase, as it entails interacting virtuously with and benefiting one's spheres of relations (e.g., the family, state).

The four *general insights* above can also be classified in terms of the inward and outward phases of self-cultivation. The first three insights arguably belong to the inward phase, as they involve the ethical formation of an individual's interior disposition. Pausing and reflecting on ethical considerations (Insight 1) involves an inward turn, wherein one examines oneself in relation to ethical ideas or considerations. This reflection can involve not just the cognitive, but also the emotional and sensory faculties (Insight 2), as is seen from Confucian methods such as Mencius' extension and sympathetic concern. Since this reflection concerns ethical considerations, it is inextricably linked to how one should treat others. People, however, might not be motivated to treat others according to Confucian ethical values, such as respect, sympathy, and care. This is especially true if they are "beginners" in self-cultivation. Confucians have thus devised strategies to "meet people where they are," that is, to make Confucian ethical goals more relatable to people. This can also be meaningfully applied to users of digital technologies (i.e., "meeting users where they are"/ Insight 3). These strategies can be incorporated in the act of reflection itself, such as in Mencian extension, wherein one recalls and appreciates a previous act of caring, in order to inspire oneself to express care towards

a new group of people. Insights 1 to 3, therefore, indicate aspects of the self-cultivation process involved in ethically training one's interior disposition, thus forming part of the first phase. Insight 4, on the other hand, belongs to the outward phase. That is, after shaping one's inner disposition according to ethical considerations, one translates this disposition into action by establishing ethical and cooperative relationships with others in one's spheres of relations. By going through this process, one is able to develop and express one's moral capacities, thereby achieving an essential component of the good life (i.e. morality).[112]

Since all four *general insights* involve digital technologies and can represent the different parts of self-cultivation, this suggests the possibility that digital technologies may assist the *entire* process of self-cultivation. Coupling this with the idea that self-cultivation is the means for attaining the good life yields a fifth *general insight*, which is a synthesis of the four previous ones. This insight suggests that *digital technologies have the potential to support the entire process of self-cultivation, through which people can attain the good life*.

This insight can encourage a more proactive and optimistic approach to pursuing the digital good life. Recall that the question of the digital good life asks: "How can people maintain or attain the good life in relation to digital technologies?" This question is motivated by the potential impact that digital technologies have on people's prospects for living well, due to the former's capacity for social and conceptual disruption. The notion of "disruption," however, might give the impression that digital technologies' potential impact on our lives will largely be negative; it might be thought that these technologies will only disturb or upend the everyday conventions that already make our lives run smoothly. Consequently, the question of how to attain a good life in relation to digital technologies might be pessimistically construed as an attempt to spare humans from the "dire and looming threat" that these technologies pose, and that the best we can do is to merely "survive" this threat by evading their adverse effects.

On the other hand, what this final insight suggests is that digital technologies can actually help us *flourish*, and that they have the potential to do so at every step of the way. Digital technologies can help us reflect on ourselves, enabling us to cultivate our moral strengths and establish ethical and healthy relationships with others, thereby enriching our lives. Thus, instead of viewing digital technologies as a potential yet ever-present hurdle to achieving a well-lived life, we can view them as an ever-at-hand

112 This is the conception of the good life assumed by this dissertation, which is elaborated in section 1.1, and which resonates with the Confucian understanding of the good life in which moral goods play a central part.

bridge that can help us reach that goal. To be sure, we must be wary of falling to the other extreme, namely, a naïvely excessive faith in digital technologies and technosolutionism. Ultimately, striking a healthy balance between optimism and caution regarding the potential of digital technologies appears to be the best way forward. This, too, would be what a Confucian-inspired approach recommends since hitting the "mean" (*zhong* 中) had always been important to Confucius.

It is worth noting that the philosopher Matthew Dennis has already insightfully suggested the potential of digital technologies for self-cultivation (2020, 2020a, 2023). However, the technologies he has in mind are limited to self-care and mindfulness apps, as well as social media. This dissertation, however, suggests that online multiplayer games can also promote self-cultivation. Moreover, Dennis' idea of how social media can be a means for self-cultivation depends on ethical exemplars and how their admirers will emulate them (2023). This dissertation explores other ways in which social media can promote self-cultivation, such as design features that encourage reflection. Notably, whereas Dennis relies on benign relations for social media's potential for self-cultivation – that is, the relations between exemplar and admirer – this dissertation also explores the potential of hostile relations for self-cultivation. Specifically, the dissertation demonstrates how online interactions between polarized individuals or groups can serve as an opportunity for developing moral or prosocial qualities, such as care and respect. A similar case applies to the Confucian-inspired approach to toxicity in online multiplayer games; the potentially toxic relations players establish with each other can be the locus for fostering respectful and sympathetic interactions. What these examples illustrate is the Confucian idea that interaction with people, regardless of who they are, can be an opportunity for self-cultivation, as such interaction can always be a learning moment for moral behavior. As Confucius says,

> When walking with two other people, I will always find a teacher among them. I focus on those who are good and seek to emulate them, and focus on those who are bad in order to be reminded of what needs to be changed in myself (*Analects* 7.22/ Confucius, 2003, p.71)

The above, then, are the five *general insights* that, together with the *specific insights* in section 6.1.1, answer the main research question. However, the overall research project has yielded other ideas and implications worth noting, although they may not directly answer the main research question. These will be discussed in the next section.

6.2. Further Ideas or Implications Drawn From the Research Project

Four additional insights or implications can be drawn from the general findings of this research, which may be worth reflecting on. One insight is noteworthy because it might not be expected based on how Confucian philosophy was described in the Introduction of this dissertation. This is the idea that (1) Confucian philosophy includes a certain kind of "individualism." Another insight is interesting because it reinforces the notion, claimed in the Introduction of this dissertation, that Confucian philosophy has a relational view of flourishing. This is the idea that (2) the Confucian-inspired approach to the digital good life yields recommendations that do not directly or immediately provide non-moral benefit for individual users. The final two implications can be considered as future areas of study, which the limitations of this research highlight. These are (3) the need for empirical research to assess the effectiveness of the Confucian-inspired recommendations and (4) the importance of looking to other non-Western philosophies to enrich the discourse on the digital good life.

Let me begin with the first insight, namely, that (1) Confucian philosophy includes a certain kind of "individualism." The Confucian-inspired approach to the question of the digital good life, as demonstrated in this dissertation, entails effectuating change within the self. This suggests the idea that Confucian philosophy, despite its focus on the relational aspects of human existence, is still interested in the human "individual." Indeed, in the *Great Learning*, the process of attaining the good life starts with the rectification of the self. Only when the self is rectified is a person's virtue able to affect their other spheres of relation. Thus, the self still has a crucial role to play, as the cultivation of virtue and the spread of its effects to other spheres of relations cannot occur without effort on the part of the self. That is why Confucian philosophy also emphasizes the importance of personal initiative and perseverance in striving to become good (Yunping, 2002).

The takeaway from this is that, even though Confucian philosophy is employed in this dissertation to counteract the "individualistic bias" that specific Western theories of well-being might exhibit, this does not mean that Confucian philosophy places no importance on the individual. Instead, what arguably distinguishes Confucian thought from the aforesaid Western theories of well-being is that the Confucian individual's life is normatively, and ultimately, oriented towards benefiting their spheres of relations (mainly through the individual's cultivation of moral virtues). In this way, the Confucian individual is inextricably intertwined with their spheres of relations. It is

for this reason that it might be said that the Confucian good life encompasses the individual and their spheres of relations.

This connects with the second insight, which suggests that (2) the Confucian-inspired approach to the digital good life yields recommendations that do not directly or immediately provide non-moral advantages for individual users. Some of the Confucian recommendations for addressing the four issues often entail cultivating other-regarding virtues that do not directly benefit the individual user, aside from helping them become more virtuous. This touches on the distinction between well-being and the good life mentioned in the Introduction, where the latter captures the idea of a "choiceworthy" life that might entail acting virtuously towards others even if doing so might not benefit oneself (in a non-moral sense). For example, in the context of toxic behavior in *League of Legends*, the recommendation of viewing one's toxicity from the perspective of the player who experienced it aims to cultivate the virtue of sympathetic understanding towards others. However, one could argue that, in the context of *League*'s competitive gameplay, being sympathetic towards enemy players does not directly benefit the user. Having sympathy for the adversary might even leave a user vulnerable to derision from their teammates because such an attitude could lessen their chances of winning. Indeed, having a generally acute sense of sympathy for others, even outside *League*'s setting, does not appear to directly benefit an individual in a non-moral way. Another arguable example is the Zhu-Xi-inspired mindfulness or ZIM. Although ZIM is designed to help users identify and address structural asymmetries in their social context, the benefits they may obtain from this are unlikely to be immediate. Users would likely need to exert a substantial amount of effort – and perhaps even bear some stress – to implement the structural changes ZIM might suggest. The takeaway from both examples is that, by offering recommendations that may not directly or immediately benefit the individual in a non-moral way, Confucian philosophy appears to conceive of the good life in a less individualistic sense. This, in turn, can also support the idea suggested in the Introduction that Confucian philosophy is distinct from the potentially more individualistic Western theories of well-being in that the subject of flourishing in the Confucian view is not just the individual but also their spheres of relation.

The limitations of this research are also worth reflecting on. This dissertation, due to its philosophical nature, has taken a conceptual approach. Although citations from empirical studies have supported some of the Confucian-inspired ideas, the recommendations presented in the dissertation have largely not been assessed empirically. Thus, the third implication is that (3) there is a need for empirical research

to assess the effectiveness of the Confucian-inspired recommendations. Some of these recommendations are arguably easy to assess empirically. For example, evaluating the Mencian prompt empirically can be done similarly to how previous empathic prompts have been assessed, such as those of Saveski et al. (2022). Certain recommendations, however, might be more difficult to assess empirically. For instance, the conceptual resilience of an "indicative concept" might be challenging to empirically ascertain since it might require longitudinal studies among different populations regarding the use of the concept. Regardless, the project of empirically assessing these recommendations does indeed point to future areas of research that connect with this dissertation.

Another limitation of this research is that, although it has suggested the importance of including non-Western philosophy in the discourse on the digital good life, it has only delved into Confucian philosophy. Other non-Western philosophical traditions, such as Daoist, Buddhist, and Indian schools of thought, have yet to be substantially explored. Indeed, this dissertation has not even exhausted the Confucian tradition. This dissertation draws on Classical and Neo-Confucian philosophy, as the author has a greater familiarity with these traditions. However, other branches or iterations of Confucian philosophy may be worth exploring. In particular, New Confucianism might be a promising area to explore because it developed closer in time to our digital age. Thus, this limitation of the research points to (4) the continuing importance of exploring other non-Western philosophies to enrich the discourse on the digital good life.

6.3. Scientific Implications and Relevance for Society: Fulfilling the Goals of ESDIT

The Ethics of Socially Disruptive Technologies (ESDIT) program is a 10-year research program participated in by seven Dutch educational institutions. In general, the program aims to develop philosophical insights, frameworks, and methods that can help address the ethical implications of socially disruptive technologies (SDTs).[113] In this way, the program intends to produce academic output that has implications and relevance for society. Since this dissertation is a part of the ESDIT program, its implications and relevance for society can be articulated in terms of how it fulfils ESDIT's goals or objectives.

[113] For the meaning of "socially disruptive technologies," please refer to the Introduction as well as the introduction section of chapter five.

This dissertation mainly addresses objectives two, three, and five:[114]

2. Examining critical philosophical and ethical concepts challenged by SDTs and investigating what reassessments, revisions, and innovations are needed, taking into account philosophical insights from non-Western traditions.

3. Developing innovative, comprehensive, and inclusive approaches in ethics and philosophy to study the new generation of SDTs and their impacts on humans, nature, and society.

5. Studying conceptual disruption and proposing revisions of philosophical concepts to innovate the field of practical philosophy and, by implication, the ethics of technology.

The dissertation addresses objective 2 by showing that SDTs, such as digital technologies, have challenged the dominant philosophical frameworks of well-being or the good life. This is because digital technologies have made human beings much more interconnected than before. This raises questions about the efficacy of the dominant philosophical frameworks of well-being, which primarily focus on individuals as the subjects of well-being, thereby risking the conceptual isolation of the individual and portraying human existence as atomistic. Conversely, these frameworks are also at risk of overlooking the relational aspects of human existence that can either contribute to or detract from the good life. To address this issue, this dissertation looks at a non-Western philosophical tradition, namely Confucianism, and its relational vision of the good life. Drawing inspiration from the Confucian classic, the *Great Learning*, the dissertation articulates a general framework for the good life where the latter supervenes not just on the individual human being but on the various spheres of relations in which they are situated (i.e., the interpersonal, political, and the "world). Such a framework is arguably better able to track the influence of digital technologies on the good life by highlighting how the relationships established by such technologies can affect users' chances of living well.

The dissertation also addresses objective 3 by "developing innovative, comprehensive, and inclusive approaches in ethics and philosophy to study the new generation of SDTs and their impacts," as well as articulating solutions to the potentially negative affordances of SDTs. This dissertation proposes innovative

[114] Taken and slightly modified from the ESDIT webpage: (Ethics of Socially Disruptive Technologies (ESDiT). (n.d.). About ESDiT. https://www.esdit.nl/about-esdit/)

solutions to such affordances by drawing inspiration from a largely untapped conceptual resource, namely, Confucian philosophy.

Finally, the dissertation fulfils objective 5 by "proposing revisions of philosophical concepts" as well as introducing new concepts. The dissertation fulfills this in chapter 5, where it introduces the idea of "conceptual resilience," along with its two modes of CRI and CRA, as a heuristic or guide for engineering concepts to be resistant to conceptual disruption. Conceptual resilience is an innovative idea, as it introduces a largely novel approach to addressing conceptual disruption, namely, by engineering concepts that cannot be disrupted in the first place. Preventing conceptual disruption, in turn, can potentially enhance ethical discourses about AI systems by ensuring that the concepts central to these discourses continue to function, thereby also ensuring that these discourses continue to help people determine how to govern AI for human safety and flourishing.

6.4. Final Reflection: Digital Technologies as "Good" Socially Disruptive Technologies

One of the metaphors used by Confucians to describe the process of self-cultivation is that of sculpting an uncarved piece of jade. Just as an uncarved piece of jade is chiseled into an elegant work of art, self-cultivation involves training one's unrefined dispositions and shaping them into something morally beautiful. The metaphor can inspire those seeking virtue by illustrating the appealing outcome of self-cultivation. However, the metaphor also indicates the challenging journey towards the goal of flourishing. The idea of sculpting or chiseling implies a disturbance of one's current state. One cannot hope to achieve the good or choiceworthy life while remaining the same. One is required to change, and some of these changes will be uncomfortable because they will compel one to go against habituated ways of thinking, feeling, and acting. In other words, self-cultivation entails disruption. This dissertation was motivated, among other things, by the socially disruptive capacity of digital technologies and their potential impact on people's chances of living well. Drawing inspiration from Confucian thought, this dissertation suggests that digital technologies, by virtue of their socially disruptive capacities, can serve as the chisel by which people today might carve out a good life – digital technologies can disrupt us *for the better*. This, however, requires a preliminary disruption, one involving the way we think about and design digital technologies. We must more purposefully and thoughtfully consider these technologies as means for self-cultivation. The chisel, as it were, must first be made. Thus, the challenge of living well together in the face of

digital technologies lies not so much in the disruptive capacities of digital technologies as in how we harness such capacities to achieve the good life.

References

"Announcer." 2020. League of Legends Wiki. Last modified May 1, 2020. https://leagueoflegends.fandom.com/wiki/Announcer.

"Champion Statistic." 2021. League of Legends Wiki. September 14, 2021. https://leagueoflegends.fandom.com/wiki/Champion_statistic.

"Global Video Game Consumer Population Passes 3 Billion." 2020. DFC Intelligence. August 14, 2020. https://www.dfcint.com/globalvideogame consumerpopulation/.

Daxue & Zhongyong (I. Johnston & P. Wang, Eds. & Trans.). (2012). The Chinese University of Hong Kong Press.

Aldbyani, A. (2025). Exploring Islamic mindfulness: Cultural practices and their impact on public health outcomes. *Mindfulness*, 16(3), 695-701.

Ames, R. (2021). *Human becomings: Theorizing persons for Confucian role ethics*. State University of New York Press.

Ames, R. T., & Hall, D. L. (Eds.). (2001). *Focusing the familiar: A translation and philosophical interpretation of the Zhongyong*. University of Hawaii Press.

Analayo, V. (2006). Mindfulness in the Pali Nikayas. In D.K. Nauriyal, M.S. Drummond, & Y.B. Lal (Eds.) *Buddhist Thought and Applied Psychological Research: Transcending the Boundaries*. (pp. 229–249). Routledge.

Anderson, C. A., et al. (2010). "Violent Video Game Effects on Aggression, Empathy, and Prosocial Behavior in Eastern and Western Countries: A MetaAnalytic Review." *Psychological Bulletin* 136 (2): 15173.

Angle, S. C. (2009). *Sagehood: The contemporary significance of Neo-Confucian philosophy*. Oxford University Press.

Angle, S. C., & Jin, Y. (Eds.). (2025). *Progressive Confucianism and its critics: Dialogues from the Confucian heartland*. Routledge.

Angle, S., & Tiwald, J. (2017). *Neo-Confucianism: A Philosophical Introduction*. Polity.

Annas, J. (1989). Self-love in Aristotle. *The Southern Journal of Philosophy*, *27*(5), 1.

Arora, S. D., Singh, G. P., Chakraborty, A., & Maity, M. (2022). Polarization and social media: A systematic review and research agenda. *Technological Forecasting and Social Change*, *183*, 121942. https://doi.org/10.1016/j.techfore.2022.121942

Ashmore. R. (2004). Word and Gesture: On Xuan-School Hermeneutics of the Analects. *Philosophy East and West*, 54 (4), 458–488. https://www.jstor.org/stable/4148015

Bail, C. (2021). *Breaking the Social Media Prism: How to Make Our Platform Less Polarizing*. Princeton University Press.

Bail, C., Argyle, L., Brown, T., Bumpus, J. Chen, H., Hunzaker, M.B.F., Lee, J. Mann, M. Merhout. F. & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bal, P. & Dóci, E. (2018). Neoliberal Ideology in Work and Organizational Psychology. *European Journal of Work and Organizational Psychology*, 27 (5), 536–548. https://doi.org/10.1080/1359432X.2018.1449108

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Beres, N.A., et al. (2021). "Don't You Know That You're Toxic: Normalisation of Toxicity in Online Gaming." In CHI '21: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, May 8-13, 2021. https://dl.acm.org/doi/10.1145/3411764.3445157

Bhatnagar, R. (2020). "Top 5 Most Toxic Video Game Communities." sportskeeda. August 24, 2020. https://www.sportskeeda.com/esports/the top5toxicvideogamecommunities.

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*, 11(3), 230–241. https://doi.org/10.1093/clipsy.bph077

Bodhi, B. (2011). What does Mindfulness Really Mean? A Canonical Perspective. *Contemporary Buddhism*, 12 (1), 19–39. doi: 10.1080/14639947.2011.564813

Bonanno, G. A. (2004). Loss, Trauma, and Human Resilience: Have We Underestimated the Human Capacity to Thrive After Extremely Aversive Events? *American Psychologist, 59*(1), 20–28. https://doi.org/10.1037/0003-066X.59.1.20

Bonenfant, M., Lakkaraju, K., Sukthankar, G., & Wigand, R. T. (2018). Toxic allies and caring friends: Social systems and behavioral norms in *League of Legends* and *Guild Wars 2*. In K. Lakkaraju, G. Sukthankar, & R. T. Wigand (Eds.), *Social interactions in virtual worlds: An interdisciplinary perspective* (pp. 130–161). Cambridge University Press.

Bosch, M., Fernandez-Borsot, G., Comas, A. M. I., & Vaello, J. F. (2022). Evolving friendship? Essential changes, from social networks to artificial companions. *Social Network Analysis and Mining*, *12*(1). https://doi.org/10.1007/s13278-022-00864-1

Bradley, B. (2015). *Well being*. Polity Press.

Brey, P. (2012). Well being in Philosophy, Psychology, and Economics. In P. A. E. Brey, A. R. Briggle, & E. H. Spence (Eds.), *The Good Life in a Technological Age.* (pp. 15–34). Routledge.

Brey, P. (2018). "The Strategic Role of Technology in a Good Society." *Technology in Society* 52: 3945. https://doi.org/10.1016/j.techsoc.2017.02.002.

Brey, P., Briggle, A., & Spence, E. H. (Eds.). (2012). *The Good Life in a Technological Age*. Routledge.

Brinks, M. (2020). "12 Games That Can't Escape Their Own Aggressively Toxic Communities." Ranker. October 16, 2020. https://www.ranker.com/list/videogameswithtoxiccommunities/melissabrinks.

Bruns, A. (2019). *Are Filter Bubbles Real*? Wiley.

Buder, J., Rabl, L., Feiks, M., Badermann, M., & Zurstiege, G. (2021). Does negatively toned language use on social media lead to attitude polarization? *Computers in Human Behavior, 116,* 106663. https://doi.org/10.1016/j.chb.2020.106663

Burr., C. Tadeo., M., & Floridi L. (2020) The Ethics of Digital Well-Being: A Thematic Review. Science and Engineering Ethics. *Science and Engineering Ethics*, 26, 2313–2343. https://doi.org/10.1007/s11948-020-00175-8

Caldeira, S., & Timmins, F. (2016). Resilience: synthesis of concept analyses and contribution to nursing classifications. *International nursing review*, *63*(2), 191-199.

Calm. (2023). *What is Mindfulness Meditation?* https://www.calm.com/blog/what-is-mindfulness-meditation

Cannon, J. (2016). 'Education as the Practice of Freedom: A Social Justice Proposal for Mindfulness Educators.' In R. Purser, D. Forbes & A. Burke (Eds.), *Handbook of Mindfulness: Culture, Context, and Social Engagement*. Springer International Publishing.

Carvalho, A., & Grácio, R. (2022). The Dark Side of Mindfulness: Workplace Socialization, Neoliberalism and the Self. *Communication & Language at Work*, 8 (2), 63-77, https://doi.org/10.7146/claw.v8i2.135105

Carroll, T. (2023). Mindfulness as praeparatio evangelica in the Irish Catholic Primary School. *Irish Theological Quarterly*, 88(4), 323-339.

Cavanna, A. E., Purpura, G., Riva, A., Nacinovich, R., & Seri, S. (2023). The Western origins of mindfulness therapy in ancient Rome. *Neurological Sciences*, 44(6), 1861-18

Chan, W. (1963). *A Sourcebook in Chinese Philosophy*. Princeton University Press.

Chems-Maarif, R., Cavanagh, K., Baer, R., Gu, J., & Strauss, C. (2025). Defining mindfulness: A review of existing definitions and suggested refinements. *Mindfulness*, 16(1), 1-20.

Chen, X. (2019). 'Family-culture' and Chinese politeness: An emancipatory pragmatic account. *Acta Linguistica Academica*, 66 (2), 251–270.

Chiesa, A. (2013). The difficulty of defining mindfulness: Current thought and critical issues. *Mindfulness*, 4(3), 255-268.

Chhabra, S., & Resnick, P. (2012). CubeThat: News Article Recommender. In *RecSys '12: Proceedings of the sixth ACM conference on Recommender systems*, (pp.295-296). Association for Computing Machinery. https://doi.org/10.1145/2365952.2366020

Ching, J. (2000). *The Religious Thought of Chu His*. Oxford University Press.

Chopra, D. (2024). *Digital dharma: How AI can elevate spiritual intelligence and personal well being*. Random House.

Cline, E. M. (2016). "The Highways and Byways of Ritual: Pascal and Xunzi on Faith, Virtue, and Religious Practice." *European Journal for Philosophy of Religion* 8 (1): 325.

Confucius. (2003). *Analects: With Sections from Traditional Commentaries*. (Slingerland, E., Trans.) Hackett.

Confucius. (2017). *Understanding the Analects of Confucius: A new translation of Lunyu with annotations* (P. Ni, Trans.). State University of New York Press.

Constantinescu, M., Uszkai, R., Vică, C., & Voinea, C. (2022). Children-Robot friendship, Moral agency, and Aristotelian Virtue Development. *Frontiers in Robotics and AI*, *9*. https://doi.org/10.3389/frobt.2022.818489

Cooper, J. M. (1985). Aristotle on the Goods of Fortune. *The Philosophical Review*, *94*(2), 173-196.

Cottine, C. (2016). Role Modeling in an Early Confucian Context. *The Journal of Value Inquiry*, *50*(4), 797–819. https://doi.org/10.1007/s10790-016-9576-3

Cretney, R. (2014). Resilience for whom? Emerging critical geographies of socio-ecological resilience. *Geography compass*, *8*(9), 627–640.

Cuhadar, D., Tanriverdi, D., Pehlivan, M., Kurnaz, G., & Alkan, S. (2016). Determination of the psychiatric symptoms and psychological resilience levels of hematopoietic stem cell transplant patients and their relatives. *European journal of cancer care*, *25*(1), 112-121.

Dai, Z. (1990). *Tai Chen on Mencius: Explorations in words and meaning* (A. Chin & M. Freeman, Trans.). Yale University Press.Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. G.P. Putnam.

Danaher, J. (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, *3*(1), 5–24. https://doi.org/10.5325/jpoststud.3.1.0005

Davies, W. (2014). *The Limits of Neoliberalism*. Sage.

de Bary, W. (1981). *Neo-Confucian Orthodoxy and the Learning of the Mind-and-Heart*. Columbia University Press.

De Graaf, M. M. A. (2016). An Ethical evaluation of Human–Robot Relationships. *International Journal of Social Robotics*, *8*(4), 589–598. https://doi.org/10.1007/s12369-016-0368-5

Dennis, M. J. (2020). Cultivating digital well-being and the rise of self-care apps. In C. Burr & L. Floridi (Eds.), *Ethics of digital well-being* (Vol. 140, pp. 119-137). Springer. https://doi.org/10.1007/978-3-030-50585-1_6

Dennis, M. (2020a). Technologies of self-cultivation. How to improve Stoic self-care apps. *Human Affairs*, *30*(4), 549–558.

Dennis, M. J. (2023). Nietzsche's Untimely Prophecy: Online Exemplars and Self-Cultivation. *Educational Theory*, *73*(5), 749-761.

Dennis, M. J., & Clancy, R. F. (2022). Intercultural ethics for digital well being: Identifying problems and exploring solutions. *Digital Society*, 1, 7. https://doi.org/10.1007/s44206-022-00006-2

Dennis, M. J., & Ziliotti, E. (2022). Living well together online: Digital well being from a Confucian perspective. *Journal of Applied Philosophy, 40*(2), 263–279. https://doi.org/10.1111/japp.12627

de Terte, I., Stephens, C., & Huddleston, L. (2014). The development of a three-part model of psychological resilience. *Stress and Health*, 30(5), 416–424. https://doi.org/10.1002/smi.2625

Dilts, A. (2011). From 'Entrepreneur of the Self' to 'Care of the Self': Neo-liberal Governmentality and Foucault's Ethics. *Foucault Studies*,12, 130-146. https://doi.org/10.22439/fs.v0i12.3338

DongHuaP. (2021). "League of Documentary The Harsh Truth about Toxic Korean Culture." Youtube. July 2, 2021.Video, 27:13. https://youtu.be/u4FXLkeOSg.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society, 21*(5), 729-745. https://doi.org/10.1080/1369118X.2018.1428656

Edenberg, E. (2021). The problem with disagreement on social media: Moral not epistemic. In E. Edenberg & M. Hannon (Eds.), *Political epistemology* (online ed). Oxford University Press. https://doi.org/10.1093/oso/9780192893338.003.0015

Eisenstadt, M., Liverpool, S., Infanti, E., Ciuvat, R.M., & Carlsson, C. (2021). Mobile Apps That Promote Emotion Regulation, Positive Mental Health, and Well-being in the General Population: Systematic Review and Meta-analysis. *JMIR Ment Health*, 8 (11), e31170. doi: 10.2196/31170.

Elder, A. M. (2017). *Friendship, robots, and social media: False friends and second selves*. Routledge.

ESB Staff. (2021). "Top 10 Most Toxic Gaming Communities in The World". Esportsbets. October 13, 2021. https://www.esportsbets.com/news/toxic gamingcommunities/.

Ess, C. (2006). Ethical pluralism and global information ethics. *Ethics and Information Technology*, *8*(4), 215-226.

Ess, C. (2008). Culture and Global Networks: Hope for a Global Ethics? In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy* (pp. 195–225). chapter, Cambridge: Cambridge University Press.

European Commission High-Level Expert Group on Artificial Intelligence (AI HLEG). (2019). Ethics guidelines for trustworthy AI. European Commission. Available Online at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*. Official Journal of the European Union, L 192, July 12, 2024, pp. 1–84.

Floridi, L. (2015). *The Online Manifesto: Being Human in a Hyperconnected Era*. London: Springer.

Forbes, D. (2016). Modes of Mindfulness. Prophetic Critique and Integral Emergence. *Mindfulness*, 7 (6), 1256–1270, https://doi.org/10.1007/s12671-016-0552-6

Forbes, D. (2019). *Mindfulness and Its Discontents: Education, Self, and Social Transformation*. Winnipeg: Fernwood Publishing.

Forbes, D. (2022). Neoliberal Hegemonic Masculinity and McMindfulness: The Need for Buddhist Values and Principles in Mindful Masculinity Programs. *Religions*, 13 (544), 1–13. https://doi.org/10.3390/rel13060544

Fröding, B., & Peterson, M. (2020). Friendly AI. *Ethics and Information Technology*, *23*(3), 207–214. https://doi.org/10.1007/s10676-020-09556-w

Furtak, R. A. (2019). *Knowing emotions: Truthfulness and recognition in affective experience*. Oxford University Press.

Garimella, K., Morales, G., Gionis A., & Mathioudakis, M. (2017). Reducing Controversy by Connecting Opposing Views. In *WSDM '17: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, (pp.81–90). Association for Computing Machinery. https://doi.org/10.1145/3018661.3018703

Gelles, H. (2015). *Mindful Work*. Houghton Mifflin Harcourt.

Gentzkow, M. (2016). *Polarization in 2016*. Toulouse Network for Information Technology Whitepaper, 1–23.

Gergen, K. (2009). *Relational Being: Beyond Self and Community*. Oxford: Oxford University Press.

Gethin, R. (2011). On Some Definitions of Mindfulness. *Contemporary Buddhism*, 12 (1), 263–279. https://doi.org/10.1080/14639947.2011.564843

Gidaris, C. (2023). The Neoliberal Perils of Yoga and Self-Care on Apps and Platforms. *International Journal of Cultural Studies*, 26 (5), 606–620. doi: 10.1177/13678779231179739.

Gill, R., & Donaghue, N. (2016). Resilience, apps and reluctant individualism: Technologies of self in the neoliberal academy. *Women's Studies International Forum*, 54, 91–99. https://doi.org/10.1016/j.wsif.2015.06.016

Grandprey-Shores, K., He, Y., Swanenburg, K.L., Kraut, R., & Riedl, J. (2014). "The Identification of Deviance and its Impact on Retention in a Multiplayer Game." In *CSCW 14: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, Baltimore*, February 1519, 2014. https://dl.acm.org/doi/pdf/ 10.1145/2531602.2531724.

Greitemeyer, T. (2022). "The Dark and Bright Side of Video Game Consumption: Effects of Violent and Prosocial Video Games." *Current Opinion in Psychology* 46 (101326):15.

Griffiths, D. (2023, December). *Strengthening the human rights system: Barometer in context*. Open Society Foundations. https://www.opensocietyfoundations.org/publications/open-society-barometer-can-democracy-deliver

Grundy, Q. (2022). A Review of the Quality and Impact of Mobile Health Apps. *Annual Review of Public Health*, 43, 117–134. https://doi.org/10.1146/annurev-publhealth-052020-103738

Gunderson, L. H. (2000). Ecological resilience in theory and application. *Annual Review of Ecology and Systematics, 31*(1), 425–439.

Harvey, D. (2005). *A brief history of neoliberalism*. Oxford University Press. https://doi.org/10.1093/oso/9780199283262.001.0001

Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs, 34*(1), 31–55. https://doi.org/10.1111/0029-4624.00201

Haslanger, S. (2012). *Resisting reality: Social construction and social critique.* Oxford University Press.

Haybron, D. M. (2008). *The pursuit of unhappiness: The elusive psychology of well being*. Oxford University Press.

Headspace. (2024) *8 more reasons to give mindfulness a shot*. https://www.headspace.com/articles/8-reasons-for-mindfulness

Henderson, J. B., & Ng, O. C. (2014). The Commentarial Tradition. In *Dao Companion to the Analects* (pp. 37-53). (Dao Companions to Chinese Philosophy; Vol. 4). Springer Science and Business Media B.V. https://doi.org/10.1007/978-94-007-7113-0_3

Higgs, E., Light, A., & Strong, D. (Eds.). (2004). *Technology and the good life?* University of Chicago Press.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics, 4*(1), 1–23.)

Hongladarom, S. (2021). Charles Ess's Pros Hen Ethical Pluralism: An Interpretation. *Journal of Contemporary Eastern Asia*, *20*(1).

Hopster, J., Brey, P., Klenk, M. B. O. T., Löhr, G., Marchiori, S., Lundgren, B., & Scharp, K. (2023). Conceptual disruption and the ethics of technology. In I. van de Poel, L. Frank, J. Hermann, J. Hopster, D. Lenzi, S. Nyholm, B. Taebi, & E. Ziliotti (Eds.), *Ethics of Socially Disruptive Technologies: An Introduction* (pp. 141–162). Open Book Publishers. https://doi.org/10.11647/obp.0366.06

Hu, J. I. (2019). Part 1: Moral motivation in Mencius —When a child falls into a well. *Philosophy Compass*, *14*(8). https://doi.org/10.1111/phc3.12615

Huberty, J., Puzia, M.E., Green, J., Vlisides-Henry, R.D., Larkey, L., Irwin, M. R., & Vranceanu, A. (2021). A mindfulness meditation mobile app improves depression and anxiety in adults with sleep disturbance: Analysis from a randomized controlled trial. *General Hospital Psychiatry,* 73, https://doiorg.tudelft.idm.oclc.org/10.1016/j.genhosppsych.2021.09.004

Hurka, T. (2013). Aristotle on virtue: Wrong, wrong, and wrong. In *Aristotelian ethics in contemporary perspective* (pp. 9-26). Routledge.

Hutton, E. L. (2002). Moral Connoisseurship in Mengzi. In Liu, X. & Ivanhoe, P.J. (eds), *Essays on the Moral Philosophy of Mengzi* (pp. 163–186). Hackett.

Hyland, T. (2016). The erosion of right livelihood: counter-educational aspects of the commodification of mindfulness practice. *Person-Centered and Existential Therapies*, 15 (3), https://doi.org/10.1080/14779757.2016.1179666

Hyland, T. (2017). McDonaldizing Spirituality: Mindfulness, Education, and Consumerism. *Journal of Transformative Education* 15 (4), 334–356. https://doi.org/10.1177/1541344617696972

Ing, M. D. K. (2012). *The Dysfunction of Ritual in Early Confucianism*. United States: Oxford University Press.

Ivanhoe, P. J. (2000). *Confucian moral self cultivation*. Hackett Publishing.

Ivanhoe, P. J. (Ed.) (2019). *Zhu Xi: Selected Writings*. Oxford University Press.

Ivanhoe, P. J. (2014). A happy symmetry: Xunzi's ecological ethic. In T. C. Kline III & J. Tiwald (Eds.), *Ritual and religion in the Xunzi* (pp. 43–60). State University of New York Press.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. https://www.annualreviews.org/content/journals/10.1146/annurev-polisci-051117-073034

Jin, H., Spence, E.H. Internet Addiction and Well being: Daoist and Stoic Reflections. *Dao* 15, 209–225 (2016). https://doi.org/10.1007/s11712-016-9488-8

Jullien, F. (2000). *Detour and Access: Strategies of Meaning in China and Greece*. United Kingdom: Zone Books.

Kabat- Zinn, J. (1994). *Wherever You Go, There You Are*. Hyperion.

Kahan, D. M. (2012). Cultural cognition as a conception of the cultural theory of risk. In S. Roeser, R. Hillerbrand, M. Peterson, & P. Sandin (Eds.), *Handbook of risk theory* (pp. 725–759). Springer. https://doi.org/10.1007/978-94-007-1433-5_28

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kaliarnta, S. (2016). Using Aristotle's theory of friendship to classify online friendships: a critical counterview. *Ethics and Information Technology*, *18*(2), 65-79.

Kamarulbahri, T. M. S. T., Raduan, N. J. N., & Ali, N. F. (2024). The concept of mindfulness in Islam. *Journal of Islamic, Social, Economics and Development*, 9(63), 94-101.

Keenan, B. C. (2011). *Neo-Confucian self-cultivation*. University of Hawaii Press.

Kelly, B. D. (2023). Mindful, mindless, or misunderstood? A critical perspective of the mindfulness concept. *Irish Journal of Psychological Medicine*, 40(3), 491-493.

Kempt, H. (2022). *Synthetic friends: A philosophy of human-machine friendship*. Palgrave Macmillan.

Keng, S., Chin, J.W.E., Mammadova, M., & Teo, I. (2022). Effects of Mobile App-Based Mindfulness Practice on Healthcare Workers: a Randomized Active Controlled Trial. *Mindfulness*, 13 (11), 2691–2704. https://doi.org/10.1007/s12671-022-01975-8

Khetjoi, S. (2019). Friend In Buddhist Perspective: The Benefit for Social Work Practice for Getting Social Welfare. *Journal of International Buddhist Studies*, *10*(2), 93–106.

Kim, R. (2020). *Confucianism and the philosophy of well being*. Routledge.

Kim, S. (2023). *Confucian constitutionalism: Dignity, rights, and democracy*. Oxford University Press.

Kim, Y., & Kim, Y. (2019). Incivility on Facebook and political polarization: The mediating role of seeking further comments and negative emotion. *Computers in Human Behavior*, 99: 219-227. https://doi.org/10.1016/j.chb.2019.05.022

Kirloskar-Steinbach, M., & Kalmanson, L. (2021). *A practical guide to world philosophies: Selves, worlds, and ways of knowing*. Bloomsbury Publishing.

Kirmayer, L. (2015). Mindfulness in cultural context. *Transcultural Psychiatry*, 52 (4), 447–469. doi: 10.1177/1363461515598949

Kołodziejska M., & Paliński, M. (2022). 'Train your mind for a healthy life.' The medicalization of mediatized mindfulness in the West. *Current Psychology*, 42, 15210-15222. https://doi.org/10.1007/s12144-022-02814-8

Kordyaka, B., Klesel, M., & Jahn, K. (2019). "Perpetrators in League of Legends: Scale Development and Validation of Toxic Behavior." In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, Grand Wilea, Hawaii, January 8-11, 2019. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1248&context=hicss52.

Kou, Y. (2020). "Toxic Behavior in Team Based Competitive Gaming: The Case of League of Legends." In CHI PLAY '20: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Online, November 2-4, 2020. https://dl.acm.org/doi/10.1145/3410404.3414243.

Kraut, R. (2007). *What is Good and Why?: The Ethics of Well-Being*. Harvard University Press.

Kristensen, M.L. (2017). Mindfulness and resonance in an era of acceleration: a critical inquiry. *Journal of Management, Spirituality & Religion*, 12 (2), 178-195, https://doi.org/10.1080/14766086.2017.1418413

Kroesen, M. (2022). Working from home during the corona-crisis is associated with higher subjective well being for women with long (pre-corona) commutes. *Transportation Research Part A: Policy and Practice, 156*, 14–23. https://doi.org/10.1016/j.tra.2021.10.025

Lahtinen, O., Aaltonen, J., Kaakinen, J., Franklin L., & Hyönä, J. (2021). The effects of app-based mindfulness practice on the well-being of university students and staff. *Current Psychology*, 42 (6), 4412-4421. https://doi.org/10.1007/s12144-021-01762-z

Lai, K. (2016). *An Introduction to Chinese Philosophy*, 2nd ed. Cambridge: Cambridge University Press. Kindle.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lee, J. K., Choi, J., Kim, C., & Kim, Y. (2014). Social Media, Network Heterogeneity, and Opinion Polarization. *Journal of Communication*, *64*(4), 702–722. https://doi.org/10.1111/jcom.12077

Legett, W. (2022). Can Mindfulness really change the world? The political character of meditative practices. *Critical Policy Studies*, 16 (3): 261–278. https://doi.org/10.1080/19460171.2021.1932541

Leung, Y. H. (2020). Zhu Xi and the debate between virtue ethicists and situationists: Virtue cultivation as a possible, practical, and necessary enterprise. In K. C. Ng & Y. Huang (Eds.), *Dao companion to Zhu Xi's philosophy* (Vol. 13, pp. 895–927). Springer.

Lewis, C., J. (2020). *Confucian Ritual and Moral Education*. London: Lexington Books. Kindle.

Li, C. (2014). *The Confucian Philosophy of Harmony*. Routledge.

Li, C. (2023). *Reshaping Confucianism: A progressive inquiry*. Oxford University Press.

Liao, Q. V., & Fu, W. T. (2013). Beyond the filter bubble: Interactive effects of perceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2359–2368). Association for Computing Machinery. https://doi.org/10.1145/2470654.2481326

Liu, X. & Ivanhoe, P.J. (Eds.) (2002). *Essays on the Moral Philosophy of Mengzi*. Hackett.

Liu, X. (2002). Mengzian Internalism. In Liu, X. & Ivanhoe, P.J. (eds), *Essays on the Moral Philosophy of Mengzi* (pp. 101–131). Hackett.

Lu, X. (2010). Rethinking Confucian friendship. *Asian Philosophy*, *20*(3), 225–245. https://doi.org/10.1080/09552367.2010.511023

Lukoff, K., Lyngs, U., Gueorguieva, S., Dillman, E.S., Hiniker, A., & Munson, S. (2020). From Ancient Contemplative Practice to the App Store: Designing a Digital Container for Mindfulness. In *DIS '20: Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 1551–1564). ACM. http://dx.doi.org/10.1145/10.1145/3357236.3395444

Lundgren, B. (2024). Undisruptable or stable concepts: Can we design concepts that can avoid conceptual disruption, normative critique, and counterexamples? *Ethics and Information Technology*, *26*(2), 33. https://doi.org/10.1007/s10676-024-09767-5

Lundgren, B., & Möller, N. (2019). Defining Information Security. *Science & Engineering Ethics*, 25, 419–441. https://doi.org/10.1007/s11948-017-9992-1

Lynch, M.P. (2021). "Political Disagreement, Arrogance, and the Pursuit of Truth."

Macioce, F. (2016). Balancing cultural pluralism and universal bioethical standards: A multiple strategy. *Medicine, Health Care and Philosophy*, 19(3), 393–402.

Manne, K. (2018). *Down girl: The logic of misogyny.* Oxford University Press.

Marchal, N. (2022). "Be Nice or Leave Me Alone": An Intergroup Perspective on Affective Polarization in Online Political Discussions. *Communication Research*, *49*(3), 009365022110425. https://doi.org/10.1177/00936502211042516

Marchiori, S., & Scharp, K. (2024). What is conceptual disruption? *Ethics and Information Technology*, 26(1), 18. https://doi.org/10.1007/s10676-024-09749-7

Markham, T. (2020). *Digital Life*. Polity.

Marin, L. (2024). Toxic online environments are what make rational persuasion become wrongful. *Philosophy & Technology, 37*(46). https://doi.org/10.1007/s13347-024-00738-8

Marin, L., & Roeser, S. (2020). Emotions and digital well-being: The rationalistic bias of social media design in online deliberations. In C. Burr & L. Floridi (Eds.), *Ethics of digital well-being: A multidisciplinary approach* (pp. 139–150). Springer. https://doi.org/10.1007/978-3-030-50585-1_9

Mattice, S. (2015). Perspectives from China: Social media and living well in a Chinese context. In B. A. Beasley & M. R. Haney (Eds.), *Social media and living well* (pp. 123–139). Lexington Books.

McCarty, N. (2019). *Polarization: What everyone needs to know®*. Oxford University Press.

Mencius (2008). *Mengzi: With Selection from Traditional Commentaries*. (B. Van Norden, Trans.) Hackett.

Mendie, J. G., & Udofia, S. N. (2020). The idea of the good life in Aristotle and Confucius. *Jurnal Sosialisasi: Jurnal Hasil Pemikiran, Penelitian, dan Pengembangan Keilmuan Sosiologi Pendidikan, 7*(2), 61–73.

Michaelson, J. (2019). *Where should I meditate?* Ten Percent Happier. https://www.tenpercent.com/meditationweeklyblog/where-should-i-meditate

Mittelstadt, B. (2017). From individual to group privacy in big data analytics. *Philosophy & Technology, 30*(4), 475–494. https://doi.org/10.1007/s13347-017-0253-7

Mossner, C., & Walter, S. Shaping Social Media Minds: Scaffolding Empathy in Digitally Mediated Interactions?. *Topoi 43*, 645–658 (2024). https://doi.org/10.1007/s11245-024-10034-x

Munn, N., & Weijers, D. (2023). Can we be friends with AI? What risks would arise from the proliferation of such friendships? *Proceedings of the International Conference on Computer Ethics: Philosophical Enquiry (CEPE)*, Chicago, IL.

Munroe, W. (2023). Echo chambers, polarization, and "Post-truth": In search of a connection. *Philosophical Psychology*, 1–32. https://doi.org/10.1080/09515089.2023.2174426

Muriel, D. & Crawford, G. (2018). *Video Games as Culture: Considering the Role and Importance of Video Games in Contemporary Society*. New York: Routledge.

Mutz, D. (2006). *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge University Press.

Muzaki, F.I. (1990). "Confucius Ideas on Popular Video Games in Indonesia." *International Journal of Scientific Research in Science, Engineering and Technology* 7 (6): 21116.

Namsithan, S., & Patipun, J. (2024). The role of Buddhist mindfulness in reshaping social media behavior and digital well being. *Journal of Modern Academic Social Science (Online)*, 1(6), 55-71. https://so19.tci-thaijo.org/index.php/J_ASS/article/view/1104

Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, 90, 298–307. https://doi.org/10.1016/j.chb.2018.07.043

Nelimarkka, M., Laaksonen, S.M., & Semaan, B. (2018). Social Media Is Polarized, Social Media Is Polarized: Towards a New Design Agenda for Mitigating Polarization. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS' 18)* (pp. 957-970). Association for Computing Machinery. https://doi.org/10.1145/3196709.3196764

Nelimarkka, M., Rancy, J.P., Grygiel, J.& Seeman, B. (2019). (Re)Design to Mitigate Political Polarization: Reflecting Habermas' ideal communication space in the United States of America and Finland. In Lampinen, A., Gergle, D., and Sharma, D. (eds.), *Proceedings of the ACM on Human-Computer Interaction*, *Vol.3*. Art. No. 141. Association for Computing Machinery. https: //doi.org/10.1145/3359243

Nelson, E.S. (2018). Confucian Relational Hermeneutics, the Emotions, and Ethical Life. In Fairfield, P. & Geniusas. S. (eds.) *Relational Hermeneutics: Essays in Comparative Philosophy* (pp.193–248). Bloomsbury.

Nilsson, H., & Kazemi, A. (2016). Reconciling and thematizing definitions of mindfulness: The big five of mindfulness. *Review of General Psychology*, 20(2), 183-193.

Ng, E. (2016). The Critique of Mindfulness and the Mindfulness of Critique: Paying Attention to the Politics of Our Selves with Foucault's Analytic of Governmentality. In R. Purser, D. Forbes & A. Burke (Eds.), *Handbook of Mindfulness: Culture, Context, and Social Engagement*. Springer International Publishing.

Ng, O.-c. (2019). Poetry, literature, textual study, and hermeneutics. In P. J. Ivanhoe (Ed.), *Zhu Xi: Selected writings* (pp. 72–92). Oxford University Press.

Nguyen, A., & Vu, H. T. (2019). Testing popular news discourse on the "echo chamber" effect: Does political polarisation occur among those relying on social media as their primary politics news source? *First Monday*, *24*(6). https://doi.org/10.5210/fm.v24i6.9632

Nguyen, C.T. (2018). Echo Chambers and Epistemic Bubbles. *Episteme*. 2020;17(2):141-161. doi:10.1017/epi.2018.32

Nielsen, K. (2021). McMindfulness in the Era of Accelerated Life. *International Review of Theoretical Psychologies*, 1 (1): 49–62.

Niksirat, K.S., Silpasuwanchai, C., Cheng, P., Ren, X. (2019). Attention Regulation Framework: Designing Self-Regulated Mindfulness Technologies. *ACM Trans. Comput.-Hum. Interact*, 26 (6): 39:1-39:44. https://doi.org/10.1145/3359593

Nordbrandt, M. (2021). Affective Polarization in the Digital age: Testing the Direction of the Relationship between Social Media and Users' Feelings for out-group Parties. *New Media & Society*, *25*(12), 146144482110443. https://doi.org/10.1177/14614448211044393

Nussbaum, M. (2011). *Creating Capabilities: The Human Development Approach*. The Belknap Press of Harvard University Press.

Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.

Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield.

O'Donnell, A. (2015). Contemplative Pedagogy and Mindfulness: Developing Creative Attention in an Age of Distraction. *Journal of Philosophy of Education*, 49 (2), 187–202, https://doi.org/10.1111/1467-9752.12136

Olberding, A. (2012). *Moral exemplars in the Analects: The good person is that*. Routledge.

Page, R. (2012). Levelling up: Playerkilling as ethical self-cultivation. *Games and Culture, 7*(3), 238–257.

Page. R. (2016). "The Dao of Space Piracy: Ethics and Chinese Modernity in Eve Online." PhD diss., University of Hawai'i at Manoa.

Parfit, D. (1984). *Reasons and persons*. Clarendon Press.

Patel, S. S., Rogers, M. B., Amlôt, R., & Rubin, G. J. (2017). What do we mean by'community resilience'? A systematic literature review of how it is defined in the literature. *PLoS currents*, *9*, ecurrents-dis.

Paul, C. A. (2018). *The Toxic Meritocracy of Video Games: Why Gaming Culture Is the Worst.* Minnesota: University of Minnesota Press.

Peng, Y., & Zhang, Z. (2022). Theory and Meditation of Confucian Mindfulness: Analysis Based on the Concept of Vigilance in Solitude in Chinese Confucianism. *Journal of Religion and Health*, 62: 1872-1883, https://doi.org/10.1007/s10943-022-01655-w

Peters, D., Calvo, R. A., & Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in Psychology, 9*, 797. https://doi.org/10.3389/fpsyg.2018.00797

Prescott, A.T., Sargent, J.D., & Hull, J.G. (2018). "Metaanalysis of the relationship between violent video game play and physical aggression over time." *Proceedings of the National Academy of Science* 115 (40): 988288.

Purser, R. E. (2019). *McMindfulness: How Mindfulness Became the New Capitalist Spirituality*. London: Repeater Books.

Read, H. (2022). When and why to empathize with political opponents. *Philosophical Studies*. https://doi.org/10.1007/s11098-022-01837-y

Reid, T. (1969). *Essays on the active powers of the human mind* (B. Brody, Introduction). The MIT Press. (Original work published 1788)

Riot Games. 2022a. "Choose Your Champion." League of Legends. https://www.leagueoflegends.com/enus/champions/. Accessed June 12, 2022.

Riot Games. 2022b. "Player Reporting Guide and FAQ." League of Legends Support. https://support-leagueoflegends.riotgames.com/hc/en-us/articles/ 201752884-Player-Reporting-Guide-and-FAQ. Accessed June 12, 2022.

Roeser, S. (2011). *Moral emotions and intuitions*. Palgrave Macmillan.

Roeser, S. (2018). *Risk, technology, and moral emotions*. Routledge.

Roeser, S., & Pesch, U. (2016). An emotional deliberation approach to risk. *Science, Technology & Human Values*, 41(2), 274–297. https://doi.org/10.1177/0162243915596237

Rogacz, D. (2021). Operating with Names: Operational Definitions in the Analects and Beyond. *Dao*, 21, 19–35. https://doi.org/10.1007/s11712-021-09813-9

Roets, L. (2025). Centering Humanity Within Digital Wellbeing: Using Our Ubuntu Values to Tell Our Story. In: van Niekerk, A., Harry, N., Coetzee, M. (eds) *Unlocking Sustainable Wellbeing in the Digital Age. Human Well-Being Research and Policy Making*. Springer, Cham. https://doi.org/10.1007/978-3-031-87616-5_16

Rosenberger R., & Verbeek, P. (2015). A Field Guide to Postphenomenology. In R. Rosenberger & P. Verbeek (Eds.), *Postphenomenological Investigations: Essays on Human-Technology Relations*. Lexington.

Ross, W. D. *The Right and the Good.* Oxford: Clarendon Press, 1930.

Ryland, H. (2021). It's Friendship, Jim, but Not as We Know It: A Degrees-of-Friendship View of Human–Robot Friendships. *Minds and Machines*, *31*(3), 377–393. https://doi.org/10.1007/s11023-021-09560-z

Santos, G., & Azhari, R. (2022). Can we save GHG emissions by working from home? *Environmental Research Communications, 4*(035007). https://doi.org/10.1088/2515-7620/ac3d3e

Sato, H. Chu Hsi's "Treatise on *Jen*". In W.-T. Chan (Ed.) *Chu-Hsi and Neo-Confucianism* (pp. 212-227). University of Hawaii Press.

Saveski, M., Gillani, N., Yuan, A., Vijayaraghavan, P., & Roy, D. (2022). Perspective-taking to reduce affective polarization on social media. *Proceedings of the International AAAI Conference on Web and Social Media, 16*(1), 885–895. https://doi.org/10.1609/icwsm.v16i1.19343

Schaberg, D., Li, W., & Durrant, S. (2016). *Zuo Tradition / Zuozhuan左傳: Commentary on the "Spring and Autumn Annals"*. Seattle: University of Washington Press.

Scharkow, M., Mangold, F., Stier, S., & Breuer, J. (2020). How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences*, 117(6), 2761–2763. https://doi.org/10.1073/pnas. 1918279117

Scharp, K. (2013). *Replacing Truth*. Oxford, England: Oxford University Press.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Lawrence Erlbaum Associates.

Seok, B. (2013). *Embodied moral psychology and Confucian philosophy*. Lexington Books.

Sellars, J. (2018). Roman stoic mindfulness: An ancient technology of the self. In *Ethics and self-cultivation* (pp. 15-29). Routledge.

Slingerland, E. (2003). *Effortless action: Wu-wei as conceptual metaphor and spiritual ideal in early China*. Oxford University Press.

Slunecko, T. & Chlouba L. (2021). Meditation in the age of its technological mimicry. A dispositif analysis of mindfulness applications. *International Review of Theoretical Psychologies*, 1 (1), 63-77. https://doi.org/10.7146/irtp.v1i1.127079

Smith, J. E. (2015). Tradition, culture, and the problem of inclusion in philosophy. *Comparative Philosophy*, *6*(2), 4.

Solomon, R. C. (1993). *The passions: Emotions and the meaning of life*. Hackett Publishing Company.

Somers, B. (2022). Mindfulness in the Context of Engaged Buddhism: A Case for Engaged Mindfulness. *Religions*, 13 (8), 746, https://doi.org/10.3390/rel13080746

Søraker, J. H. (2008). The moral status of information and information technologies: a relational theory of moral status. In *Information Security and Ethics: Concepts, Methodologies, Tools, and Applications* (pp. 3829-3847). IGI Global Scientific Publishing.

Sovrani, D. (2022). *The many facets of mindfulness as a tool to manage work-related stress: From the implications under neoliberalism to the struggle of "being" in a capitalistic world* [Master's Thesis, Università Ca' Foscari Venezia]. Archivio istituzionale ad accesso aperto. http://dspace.unive.it/handle/10579/22433

Spezzy. (2022). "How Many People Play League of Legends?—League of Legends Player Count in 2022 (March)." LeagueFeed. March 5, 2022. https://leaguefeed.net/didyouknowtotalleagueoflegendsplayercountupdated/.

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure

Sta. Maria, J. (2020) The Philosophy of Mencius as a Way of Life: A Rapport between Mencian Confucianism and Pierre Hadot's Conception of Philosophy. *Journal of Confucian Philosophy and Culture*, *34*, 43–72.

Steinert, S., Lavinia, M., & Roeser, S. (2022). Feeling and thinking on social media: Emotions, affective scaffolding and critical thinking. *Inquiry*, 1–28. https://doi.org/10.1080/0020174X.2022.2081327

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & behavior*, 7 (3), 321–326.

Sun, J. (2014). Mindfulness in context: A historical discourse analysis. *Contemporary Buddhism*, 15(2), 394-415.

Sunstein, C. R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

Swanton, C. (2003). *Virtue ethics: A pluralistic view*. Oxford University Press.

Talisse, R.B. (2019). *Overdoing Democracy: Why We Must Put Politics in its Place*. Oxford University Press

Tan, C. (2019). Rethinking the Concept of Mindfulness: A Neo-Confucian-inspired approach. *Journal of Philosophy of Education*, 53 (2): 359–373. https://doi.org/10.1111/1467-9752.12343

Tan, S.-H. (2003). *Confucian democracy: A Deweyan reconstruction.* State University of New York Press.

Tan, S. (2019). "AI and the Confucian Conception of the Human Person: Some Preliminary Reflections." Paper presented at *AI and Human Person: Chinese and Western Perspectives, Santa Clara University, Santa Clara, CA, April 4-5, 2019.* https://www.researchgate.net/publication/345706982_AI_and_the_Confucian_Conception_of_the_Human_Person_Some_Preliminary_Reflections.

Thompson, K. (2017). "Relational Self in Classical Confucianism: Lessons from Confucius' Analects." *Philosophy East and West* 67 (3): 887907

Tillotson, J., & Foster, N. (2013). *Text, cases and materials on European Union law*. Routledge-Cavendish.

Tillman, H. C. (2019). Heaven, ghosts and spirits, and ritual. In P. J. Ivanhoe (Ed.), Zhu Xi: Selected writings (pp. 116–136). New York, NY: Oxford University Press.

Titmuss, C. (2013, July 19). The Buddha of Mindfulness. A Stress Destruction Programme. *Christopher Titmuss - The Buddha Wallah*. https://www.christophertitmussblog.org/the-buddha-of-mindfulness-the-politics-of-mindfulness

Tiwald, J. (2010). Dai Zhen on human nature and moral cultivation. In J. Makeham (Ed.), *Dao companion to Neo-Confucian philosophy* (Vol. 1, pp. 399–422). Springer. https://doi.org/10.1007/978-90-481-2930-0_19

Tiwald, J. (2010a). Dai Zhen 戴震 on Sympathetic Concern. *Journal of Chinese Philosophy*, *37*(1), 76–89. https://doi.org/10.1111/j.1540-6253.2009.01565.x

Tiwald, J. (2016). Zhu Xi's critique of Buddhism: Selfishness, Salvation, and Self-Cultivation. In J. Makeham (Ed.), *The Buddhist Roots of Zhu Xi's Philosophical Thought*. Oxford University Press.

Törnberg, P., Andersson, C., Lindgren, K., & Banisch, S. (2021). Modeling the emergence of affective polarization in the social media society. *PloS one*, 16(10), e0258259. https://doi.org/10.1371/journal.pone.0258259

Trammel, R. C. (2017). Tracing the roots of mindfulness: Transcendence in Buddhism and Christianity. *Journal of Religion & Spirituality in Social Work: Social Thought*, 36(3), 367-383.

United Nations. (1993, April 2). *Final Declaration of the Regional Meeting for Asia of the World Conference on Human Rights (The Bangkok Declaration)*. United Nations. https://digitallibrary.un.org/record/167021?ln=en&v=pdf

Uthaphan, P. & Phooriko, P. (2024). Cultivating digital mindfulness: Buddhist practices for navigating social media and enhancing well being. *Journal of Modern Academic Social Science (Online)*, 1(6),72–89. https://so19.tci-thaijo.org/index.php/J_ASS/article/view/1105

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

Van de Poel, I., Frank, L., Hermann, J., Hopster, J., Lenzi, D., Nyholm, S., Taebi, B., & Ziliotti, E. (Eds.). (2023). *Ethics of socially disruptive technologies: An introduction.* Open Book Publishers. https://doi.org/10.11647/OBP.0366

Van Norden, B. W. (2017). *Taking back philosophy: A multicultural manifesto*. Columbia University Press.

Vars. (2021). "Why League of Legends is SO TOXIC/League of Legends." YouTube. June 3, 2021. Video, 15:51. https://www.youtube.com/watch?v= 2VjLXObhtBc&t=618s.

Verbeek, P. (2011). *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226852904.001.0001

Virag, C. (2017). *The Emotions in Early Chinese Philosophy*. Oxford University Press.

Wachs, S., Wright, M. F., & Vazsonyi, A. T. (2019). Understanding the overlap between cyberbullying and cyberhate perpetration: Moderating effects of toxic online disinhibition. *Criminal Behaviour and Mental Health*, 29(3), 179-188.

Walsh, Z. (2016). A Meta-Critique of Mindfulness Critiques: From McMindfulness to Critical Mindfulness'. In R. Purser, D. Forbes & A. Burke (Eds.), *Handbook of Mindfulness: Culture, Context, and Social Engagement*. Springer International Publishing.

Walsh, Z. (2018). Mindfulness under neoliberal governmentality: critiquing the operation of biopower in corporate mindfulness and constructing queer alternatives. *Journal of Management, Spirituality, & Religion*, 15 (2), 109–122. https://doi-org.tudelft.idm.oclc.org/10.1080/14766086.2017.1423239

Weeks, B.E., Lane, D.S., Kim, D.H., Lee, S.S., & Kwak., N. (2017). Incidental Exposure, Selective Exposure, and Political Information Sharing: Integrating Online Exposure Patterns and Expression on Social Media. *Journal of Computer-Mediated Communication*, 22: 363-379. https://doi.org/10.1111/jcc4.12199

West, M. (2023). *An ed-tech tragedy? Educational technologies and school closures in the time of COVID-19* (UNESCO). UNESCO. https://doi.org/10.54675/LYGF2153.

Williams, Tom, et al. (2020). "The Confucian Matador: Three Defenses Against the Mechanical Bull." In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, New York, 2020*, 2533. New York: Association for Computing Machinery.

Wilson, J. A. (2018). *Neoliberalism*. Routledge.

Wong, B. (2025). Careful, patient, and modest citizens: Facilitating civic education through Zhu Xi's method of deep reading. *Educational Philosophy and Theory*, 57(4), 374-384.

Wong, D. B. (1991). Is There a Distinction between Reason and Emotion in Mencius? *Philosophy East and West*, *41*(1), 31–44. https://doi.org/10.2307/1399716

Wong, D.B. (2015). Growing Virtue: The Theory and Science of Developing Compassion from a Mencian Perspective. In B. Burya (Ed.), *The Philosophical Challenge from China.* (pp. 23–57). The MIT Press.

Wong, D.B. (2017). Moral Sentimentalism in Early Confucian Thought. In R. Debes & K.R. Stueber (Eds.), *Ethical Sentimentalism: New Perspectives.* (pp. 230–249). Cambridge University Press.

Wong, D.B. (2023). Feeling, Reflection, and Reasoning in the *Mencius*. In Y. Xiao & K. Chong (Eds.), *Dao Companion to the Philosophy of Mencius*. (pp. 517–538). Springer.

Wong, P.-H. (2010). The 'Good Life'in Intercultural Information Ethics: A New Agenda. *The International Review of Information Ethics*, *13*, 26-32.

Wong, P.-H. (2012). Dao, harmony and personhood: Towards a Confucian ethics of technology. *Philosophy & technology*, *25*(1), 67-86.

Wong, P.-H. (2013). "Confucian Social Media: An Oxymoron?" *Dao* 12 (3): 28396.

Wong, P.-H. (2020). Cultural differences as excuses? Human rights and cultural values in global ethics and governance of AI. *Philosophy & Technology*, *33*(4), 705-715.

Wong, P.-H. (2021). Artificial intelligence, personal decisions, consent, and the Confucian idea of oneness. In P.-H. Wong & T. X. Wang (Eds.), *Harmonious technology: A Confucian ethics of technology* (pp. 79–94). Routledge.

Wong, P.-H., & Wang, T. X. (Eds.). (2021). *Harmonious technology: A Confucian ethics of technology*. Routledge.

Xiao X. (2008). Ancient Greek and Chinese Patterns of Definition: A Comparative Study. *Intercultural Communication Studies*, 7 (2), 61–77.

Xunzi. (2014). *Xunzi: The complete text* (E. Hutton, Trans.). Princeton University Press.

Yunping, W. (2002). Autonomy and the Confucian moral person. *Journal of Chinese Philosophy*, *29*(2), 251–268.

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

Zaki, J. (2019). *The War for Kindness: Building Empathy in a Fractured World*.

Zeitlin-Wu, L. (2023, November 1). Meditation apps and the unbearable whiteness of wellness. Just Tech. *Social Science Research Council*. https://doi.org/10.35650/JT.3062.d.2023

Zhu Xi. (1990). *Learning to be a Sage: Selections from the Conversations of Master Chu, arranged topically.* (Gardner, D.K., Trans.) University of California Press.

Zhu Xi. (2019). Moral Psychology and Cultivating the Self (C. Virag, Trans.) In P.J. Ivanhoe (Ed.), *Zhu Xi: Selected Writings*. (pp. 35–55). Oxford University Press.

Zhu, B., Hedman, A., & Li, H. (2017). Designing Digital Mindfulness: Presence-In and Presence-With versus Presence-Through. In *CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2685–2695). ACM. https://doi.org/10.1145/3025453.3025590

Zhu, Q., Williams, T., & Wen, R. (2019). Confucian robot ethics. *Computer Ethics-Philosophical Enquiry (CEPE) Proceedings*, *2019*(1), 12.

Zhu, Q. (2023). Confucian ethics of technology. In *Technology Ethics* (pp. 93–101). Routledge.

Ziliotti, E. (2022). *A Confucian Perspective on Polarising Technologies and their Impact on the Epistemic Quality of Democratic Processes*. [PowerPoint slides]. Unpublished manuscript.

Ziliotti, E., Benavides, P., Gwagwa, A., and Dennis, M. (2023). Social Media and Democracy. In I. van de Poel, L. Frank, J. Hermann, J. Hopster, D. Lenzi, S. Nyholm, B. Taebi, & E. Ziliotti (Eds.), *Ethics of Socially Disruptive Technologies: An Introduction* (pp. 33-52). OpenBook Publishers. https://doi.org/10.11647/OBP.0366

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review, 5*(1). https://doi.org/10.14763/2016.1.401

Curriculum Vitae

Joseph Sta. Maria (1988) was born in Manila, Philippines. Joseph completed his PhD in Ethics of Technology at Delft University of Technology between October 2021 and September 2025. He acquired his M.A. in Philosophy at the Ateneo de Manila University in 2018. At the same university, he attained his B.A. in Philosophy and B.A. in Development Studies in 2011 and 2012, respectively. Joseph also served as a lecturer at the Ateneo De Manila University from 2018 to 2021. He taught the courses "Philosophy of the Human Person" and "Philosophy of Religion." Additionally, during his time as a PhD candidate, he served as a teaching assistant for various Master-level ethics courses at TU Delft, including Ethics and Engineering, Water Ethics, Philosophy of Engineering, Science and Design, and Technological Expertise. As a PhD candidate, he also presented papers that would become part of his dissertation at various philosophical conferences, such as the FPET (Forum on Philosophy, Engineering, and Technology) conference held at TU Delft in 2023, and the ISEAP (International Society of East Asian Philosophy) conference held at the University of Edinburgh, also in 2023. Joseph's philosophical areas of interest are Classical Chinese philosophy, Western Medieval Philosophy, Comparative Philosophy, Philosophy of Religion, and Philosophy of Technology.

List of Publications

Connected to the Dissertation

1. Sta. Maria, J., & Ziliotti, E. (2022). Addressing online gaming toxicity from a Confucian perspective. *Journal of Confucian Philosophy and Culture, 38*, 131–152. https://doi.org/10.22916/jcpc.2022..38.131

Connected to the Dissertation, Submitted or Under Peer-Review

2. Sta. Maria, J., & Dennis, M. (2025). *Counteracting digital McMindfulness from a Neo-Confucian perspective*. Manuscript under external review for inclusion in the anthology *The Future of Digital Well-Being*.

Not Connected to the Dissertation

1. Sta. Maria, J. (2020). The philosophy of Mencius as a way of life: A rapport between Mencian Confucianism and Pierre Hadot's conception of philosophy. *Journal of Confucian Philosophy and Culture, 34*, 43–72. https://doi.org/10.22916/jcpc.2020..34.43
2. Sta. Maria, J. E. D. (2018). Sustaining temporal peace: A worldview for sustainable living inspired by Saint Augustine's philosophy. *Budhi: A Journal of Ideas and Culture, 22*(3), 57–99. http://dx.doi.org/10.13185/BU
3. Sta. Maria, J. E. D. (2018). The Dao admits of no admixture: Mysticism and moral realism in Zhuangzi's writings. *Budhi: A Journal of Ideas and Culture, 22*(1), 43–84.
4. Sta. Maria, J. E. D. (2018). A Zhuangzian ethic of openness and hospitality for contemporary Filipino society. *Ateneo Chinese Studies Program Lecture Series, 5*, 46–77. https://journals.ateneo.edu/ojs/index.php/cspls/issue/view/273
5. Sta. Maria, J. E. (2017). Human nature's self-revealing: A rapport between Heideggerian *physis* and *techne*, and Classical Chinese philosophy. In G. Thonhauser (Ed.), *Perspektiven mit Heidegger* (pp. 313–325). Verlag Karl Alber. https://www.herder.de/philosophie-ethik-shop/perspektiven-mit-heidegger-kartonierte-ausgabe/c-27/p-10757/
6. Sta. Maria, J. E. D. (2017). Acting without regarding: Daoist self-cultivation as education for non-dichotomous thinking. *Educational Philosophy and Theory, 49*(12), 1216–1224. https://doi.org/10.1080/00131857.2017.1364157

7. Sta. Maria, J. E. D. (2017). *Shu* and *Zhong* as the virtue of the Golden Rule: A Confucian contribution to contemporary virtue ethics. *Asian Philosophy, 27*(2), 100–111. https://doi.org/10.1080/09552367.2017.1318530

Acknowledgments

I would like to thank, first of all, my promotor, Sabine, and supervisor, Elena, for their amazing and irreplaceable mentorship throughout my time as a PhD candidate. I consider myself truly blessed to have received your insightful, kind, and supportive guidance, particularly during times when I felt uncertain and discouraged. Every time we have had meetings, I have always come out of them having greater clarity on how to proceed with my research and the motivation to continue on with it. I do not think I can repay you enough, and I can only hope that life will be as good to you as you have been good to me.

I would also like to thank the co-authors of the papers that comprise this dissertation. Without you, I would not have been able to complete my research. Thus, my gratitude extends again to Sabine and Elena, my gracious mentors who have co-authored two out of the four papers of this research. My thanks also go to Matthew and Sam, who co-authored the remaining two papers. Again, without your ideas and efforts, my project would have remained unfinished. Once more, I can only hope that life will help you accomplish your own endeavors as much as you have helped me with mine.

My thanks also go to the people I have met in the course of my PhD, whom I consider friends. Thank you for making my PhD life easier by supporting me in the various necessities that such a life requires. But, of course, what distinguishes friends from mere supporters is that the former don't just help you with your needs but also make life more *enjoyable*. So, thank you, my friends, for making this part of my life just more fun in general. I will list you here in no apparent order since, as friends, I don't think you would really mind. Also, this list is not exhaustive – there are just too many of you. So, if you think you should be on the list, consider yourself already included. Anyway, thank you, Miao, Karen, Donald, Jane, Ana, Sam, Syafira, Nick, Bahare, Sing, Yuna, Silvia, Lula, Shibu, Tony, Gladys, and everyone else who has befriended me. My heartfelt gratitude especially goes to Miao for being a virtual sister to me for three-fourths of my PhD experience, and for generously creating the cover image of this dissertation. I wish you all the best in life, and I hope we can continue to enjoy each other's company even after my PhD.

To my family who has supported me throughout my entire journey here in the Netherlands, even though they could not always be physically present, I also wish to express my deepest gratitude. Thank you to my parents, Mel and Ampy, who, through

their love, wisdom, and sacrifices, have shaped me into the person I am today and have made it possible for me to even be in this PhD program. Thank you to my two sisters, Patty and Therese, whose company, as well as successes in life, have provided me with encouragement in times of despondency. The thought that you are all living well is what I have drawn comfort from in my times of difficulty and solitude. Thus, to you, my family, I dedicate this work.

Finally, my thanks go to what the Chinese Philosopher Wang Bi designates as the ineffable "that-by-which" (*suo yi* 所以) all have their being – or the One that it points to.

The Simon Stevin Series in Ethics of Technology is an initiative of the 4TU Centre for Ethics and Technology. 4TU.Ethics is a collaboration between Delft University of Technology, Eindhoven University of Technology, University of Twente, and Wageningen University & Research. Contact: info@ethicsandtechnology.eu

Books and Dissertations

Volume 1: Lotte Asveld, '*Respect for Autonomy and Technology Risks*', 2008

Volume 2: Mechteld-Hanna Derksen, *'Engineering Flesh, Towards Professional Responsibility for 'Lived Bodies' in Tissue Engineering',* 2008

Volume 3: Govert Valkenburg, *'Politics by All Means. An Enquiry into Technological Liberalism'*, 2009

Volume 4: Noëmi Manders-Huits, *'Designing for Moral Identity in Information Technology'*, 2010

Volume 5: Behnam Taebi, *'Nuclear Power and Justice between Generations. A Moral Analysis of Fuel Cycles'*, 2010

Volume 6: Daan Schuurbiers, *'Social Responsibility in Research Practice. Engaging Applied Scientists with the Socio-Ethical Context of their Work'*, 2010

Volume 7: Neelke Doorn, *'Moral Responsibility in R&D Networks. A Procedural Approach to Distributing Responsibilities'*, 2011

Volume 8: Ilse Oosterlaken, *'Taking a Capability Approach to Technology and Its Design. A Philosophical Exploration'*, 2013

Volume 9: Christine van Burken, *'Moral Decision Making in Network Enabled Operations'*, 2014

Volume 10: Faridun F. Sattarov, *'Technology and Power in a Globalising World, A Political Philosophical Analysis'*, 2015

Volume 11: Gwendolyn Bax, *'Safety in large-scale Socio-technological systems. Insights gained from a series of military system studies'*, 2016

Volume 12: Zoë Houda Robaey, *'Seeding Moral Responsibility in Ownership. How to Deal with Uncertain Risks of GMOs'*, 2016

Volume 13: Shannon Lydia Spruit, *'Managing the uncertain risks of nanoparticles. Aligning responsibility and relationships'*, 2017

Volume 14: Jan Peter Bergen, *Reflections on the Reversibility of Nuclear Energy Technologies*, 2017

Volume 15: Jilles Smids, *Persuasive Technology, Allocation of Control, and Mobility: An Ethical Analysis*, 2018

Volume 16: Taylor William Stone, *Designing for Darkness: Urban Nighttime Lighting and Environmental Values*, 2019

Volume 17: Cornelis Antonie Zweistra, *Closing the Empathy Gap: Technology, Ethics, and the Other*, 2019

Volume 18: Ching Hung, *Design for Green: Ethics and Politics for Behavior-Steering Technology*, 2019

Volume 19: Marjolein Lanzing, *The Transparent Self: a Normative Investigation of Changing Selves and Relationships in the Age of the Quantified Self*, 2019

Volume 20: Koen Bruynseels, *Responsible Innovation in Data-Driven Biotechnology*, 2021

Volume 21: Naomi Jacobs, *Values and Capabilities: Ethics by Design for Vulnerable People*, 2021

Volume 22: Melis Baş, *Technological Mediation of Politics. An Arendtian Critique of Political Philosophy of Technology*, 2022

Volume 23: Mandi Astola, *Collective Virtues. A Response to Mandevillian Morality*, 2022

Volume 24: Karolina Kudlek, *The Ethical Analysis of Moral Bioenhancement. Theoretical and Normative Perspectives*, 2022

Volume 25: Chirag Arora, *Responsibilities in a Datafied Health Environment*, 2022

Volume 26: Agata Gurzawska, *Responsible Innovation in Business. A Framework and Strategic Proposal*, 2023

Volume 27: Rosalie Anne Waelen, *The Power of Computer Vision. A Critical Analysis*, 2023

Volume 28: José Carlos Cañizares Gaztelu, *Normativity and Justice in Resilience Strategies*, 2023

Volume 29: Martijn Wiarda, *Responsible Innovation for Wicked Societal Challenges: An Exploration of Strengths and Limitations*, 2023

Volume 30: Leon Walter Sebastian Rossmaier, *mHealth Apps and Structural Injustice*, 2024

Volume 31: Haleh Asgarinia, *Privacy and Machine Learning-Based Artificial Intelligence: Philosophical, Legal, and Technical Investigations*, 2024

Volume 32: Caroline Bollen, *Empathy 2.0: What it means to be empathetic in a diverse and digital world*, 2024

Volume 33: Iris Loosman, *Rethinking Informed Consent in mHealth*, 2024

Volume 34: Benjamin Hofbauer, *Governing Prometheus. Ethical Reflections On Risk & Uncertainty In Solar Climate Engineering Research*, 2024

Volume 35: Madelaine Ley, *It's not (just) about the robots: care and carelessness across an automated supply chain*, 2024

Volume 36: Arthur Gwagwa, *Re-imagining African Unity in a Digitally Interdependent World*, 2024

Volume 37: Jonne Maas, *Freedom in the Digital Age: Designing for Non-Domination*, 2025

Volume 38: Nynke van Uffelen, *Reconceptualising Energy Justice in light of Normative Uncertainties*, 2025

Volume 39: Cindy Friedman, *The Ethics of Humanoid Robots*, 2025

Volume 40: Tom Hannes, *What do We Call the World? A Plea for Developing an Anthropocene Morality Based on a Non-Axial Rereading of Buddhism*, 2025

Volume 41: Joost Mollen, *Prototype Ethics. Foundations for the Research Ethics of Real-World Technology Research*, 2025

Volume 42: Roxanne van der Puil, *Will democracy survive social media? Designing and regulating social media for democracy*, 2025

Volume 43: Shaked Spier, *Rethinking Platform Technologies. Moral Values, Politics, and Radical Technologies*, 2025

Volume 44: Joseph Sta. Maria, *Confucian Philosophy for the Digital Good Life*, 2025

Simon Stevin (1548-1620)

'Wonder en is gheen Wonder'

This series in the philosophy and ethics of technology is named after the Dutch / Flemish natural philosopher, scientist, and engineer Simon Stevin. He was an extraordinarily versatile person. He published, among other things, on arithmetic, accounting, geometry, mechanics, hydrostatics, astronomy, theory of measurement, civil engineering, the theory of music, and civil citizenship. He wrote the very first treatise on logic in Dutch, which he considered to be a superior language for scientific purposes. The relation between theory and practice is a main topic in his work. In addition to his theoretical publications, he held a large number of patents, and was actively involved as an engineer in the building of windmills, harbours, and fortifications for the Dutch prince Maurits. He is famous for having constructed large sailing carriages.

Little is known about his personal life. He was probably born in 1548 in Bruges (Flanders) and went to Leiden in 1581, where he took up his studies at the university two years later. His work was published between 1581 and 1617. He was an early defender of the Copernican worldview. He died in 1620, but the exact date and the place of his burial are unknown. Philosophically, he was a pragmatic rationalist. For him, wonder about a phenomenon, however mysterious, should be the starting point for seeking understanding or even ultimate explanation through human reasoning. Hence the dictum 'Wonder is no Wonder' that he used on the cover of several of his books.

This dissertation explores how drawing inspiration from Confucian philosophy can inform the pursuit of the "digital good life" or living well in relation to digital technologies. It addresses the challenges posed by technologies like personal apps, online gaming, and social media, which disrupt traditional social practices. While many Western well-being theories focus on individual flourishing, they often overlook the relational aspects of human life, especially in a digitally interconnected world. Confucian philosophy emphasizes the relational nature of the good life, inspiring a framework that considers the digital good life across four spheres: the self, interpersonal relationships, politics, and the world.

The study examines four case studies: McMindfulness in mindfulness apps (self), toxic behavior in online multiplayer games (interpersonal), political polarization on social media (political), and AI-related conceptual disruption (world). By applying Confucian insights, the research provides specific recommendations for promoting relational harmony and broader lessons on reflection, emotional and cognitive development, and shared responsibility. Ultimately, it demonstrates that digital technologies can facilitate moral self-cultivation and contribute to the good life in a connected world while enriching philosophical discourse with underrepresented non-Western perspectives.

4TU. Centre for Ethics and Technology

esdit ethics of socially disruptive technologies

TU Delft Technische Universiteit Delft

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF TWENTE.

WAGENINGEN UNIVERSITY & RESEARCH