# HOSTILE SCAFFOLDING

DISSERTATION SUBMITTED FOR THE DEGREE OF

MASTER OF ARTS IN PHILOSOPHY

AT THE

UNIVERSITY OF KWAZULU-NATAL

UNIVERSITY OF KWAZULU-NATAL™

INYUVESI YAKWAZULU-NATALI

Ryan Gary Timms

Supervisor: Prof. David Spurrett

December 2022

## Plagiarism Declaration

I, Ryan Gary Timms (Student Number: 22110845), declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a. Their words have been re-written but the general information attributed to them has been referenced

b. Where their exact words have been used, then their writing has been placed inside quotation marks and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Student Signature:

# Acknowledgments

I would like to thank everyone who has supported me throughout the writing of this dissertation. I am deeply indebted to my supervisor, Prof. David Spurrett, for all of his help and guidance. He went above and beyond throughout the supervision process and never failed to steer me in the right directions. I would also like to express my gratitude to my friends, family and loved ones. Most importantly, I would like to thank my mother and my girlfriend for their incredible love and support. I could not have undertaken this journey without them.

## Abstract

Cognitive scaffolding refers to external structures that change the cognitive demands of tasks. This dissertation begins by reviewing the literature around scaffolding—discussing key thinkers such as Hutchins (1995a), Clark (1997) and Sterelny (2003, 2010). I then develop the above scaffolding characterisation by drawing a distinction between shallow and deep scaffolding. Shallow scaffolding primarily involves cues, whereas deep scaffolding involves the significant offloading of cognitive work. By appealing to the complementarity thesis, I show that deep scaffolds can be explained through a model of cognitive extension (Menary, 2006; Sutton, 2010). Most crucially, despite the abundance of benign/ neutral examples in the literature, I argue that scaffolding can also be 'hostile' to agents' interests—benefitting one agent while undermining another. I draw on Sterelny (2003) to clarify my use of the word 'hostile' and also specify what I mean by 'interests'. I then review authors with similar concerns to my own—i.e., the lack of discussion around external structures that negatively impact agents (Aagaard, 2020; Liao & Huebner, 2020; Slaby, 2016). Despite these authors' concerns, the idea of 'hostile scaffolding' is not quite reached. I then present examples of shallow hostile scaffolding by reviewing how sunglasses (Viola, 2022) and casino interior design and ambience (Friedman, 2000; Schüll, 2012) can be hostile. Most importantly, I show that deep hostile scaffolding is a genuine concern (and not only a theoretical one) by reviewing gambling machines that use player tracking systems and virtual reel mapping (Schüll, 2012), as well as Twitter's use of gamification (Nguyen, 2021). I close the dissertation by discussing avenues for future hostile scaffolding research. This includes the ethical implications deep hostile scaffolding, scaffolding's role in forming addictive behavior, other instance of gamification, 'racist scaffolding', and hostile scaffolding in developing technologies (such as VR, dark patterns and AI).

## List of figures

**Figure 1**: Educational illustration of virtual reel mapping by Game Planit, Inc. (From Schüll, 2012, p. 88)

# Table of Contents

**Plagiarism Declaration** ii

**Acknowledgments** iii

**Abstract** iv

**List of figures** v

**Table of Contents** vi

**Introduction** 1

**Chapter 1: What is Cognitive Scaffolding?** 4

1.1. Introduction 4

1.2. The Development of 'Scaffolding': From Soviet Psychology to Cognitive Science 5

Scaffolding in Soviet Psychology & Neurophysiology 5

Scaffolding in Anglophone Psychology 6

Scaffolding's Move to the Cognitive Sciences 7

1.3. Scaffolding in Philosophy & Cognitive Science 7

Hutchins & Distributed Cognition 8

Andy Clark's Account of Scaffolding 9

Kim Sterelny's Account of Scaffolding 12

Affective Scaffolding 14

1.4. Deep & Shallow Scaffolding 16

1.5. Deep Scaffolding: How Is It Possible? 19

4E Cognition & Functionalism 20

The Parity Principle & Extended Functionalism 21

Cognitive Integration through Complementarity 23

1.6. Sterelny's Dimensions of Scaffolding 30

1.7. Conclusion: Turning Hostile 33

**Chapter 2: Making the Case for Hostile Scaffolding** 36

2.1. Introduction ........ 36
2.2. Characterising 'Hostility' ........ 37
2.3. Interests: "Who Suffers?" & "Who Benefits?" ........ 40
2.4. Humans & Scaffolding: Reliance & Vulnerability ........ 43
2.5. Other Related Work: 4E Cognition & Environmental Harms ........ 48
The Dogma of Harmony ........ 48
Mind Invasion ........ 50
Oppressive Things ........ 53
2.6. Conclusion ........ 55
**Chapter 3: Hostile Scaffolding in the Wild** ........ 56
3.1. Introduction ........ 56
3.2. Sunglasses as Hostile Scaffolding ........ 56
3.3. Hostility & Scaffolding in Las Vegas ........ 58
3.3.1. Shallow Hostile Scaffolding in Vegas ........ 59
Casino Interior Design ........ 59
Casino Ambience & Atmosphere ........ 63
3.3.2. Deep Hostile Scaffolding in Vegas ........ 67
Player Tracking & Interventions ........ 67
Slot Machines & Virtual Reel Mapping ........ 70
3.4. Deep Hostile Scaffolding & Gamification ........ 75
3.5. Conclusion ........ 77
**Chapter 4: Future Research & Conclusion** ........ 78
4.1. Future Research ........ 78
The Ethical Consequences of Deep Hostile Scaffolding ........ 78
Addiction & Scaffolding ........ 80
Racist Scaffolding ........ 81
Gamification in the Workplace (Uber, Disney & Amazon) ........ 81
Virtual Reality, Dark Patterns & AI ........ 82

4.2. Conclusion.......... 85

**Reference List** .......... 87

# Introduction

> Before the day is done, however, I hope to convince you at least of this: that the old puzzle, the mind-body problem, really involves a hidden third party. It is the mind-body-scaffolding problem. It is the problem of understanding how human thought and reason is born out of looping interactions between material brains, material bodies, and complex cultural and technological environments. We create these supportive environments, but they create us too. We exist, as the thinking things we are, only thanks to a baffling dance of brains, bodies, and cultural and technological scaffolding. (Clark, 2003, p. 11)

Over and above the standard players of mind and body, we also need to take seriously a third component: scaffolding. In the field of cognitive science, scaffolding refers to external structures that play important roles in how agents think and interact with the world. We often take for granted the extent to which scaffolding assists us in the completion of everyday tasks, both great and small. For example, road signs and lane markings indicate when to yield and stop as well as the location of traffic lanes, road centres and road edges. Their absence could compromise our driving ability and more accidents would likely occur. Similarly, our public spaces are also heavily scaffolded. Direction boards, bathroom signs and emergency exit signs act as cognitive scaffolding when they allow us to more easily locate our desired destinations, facilities and escape routes. Scaffolding is not only prevalent in the completion of small, everyday tasks but some of our most cognitively demanding accomplishments as well. Edwin Hutchins (1995a) famously put forward the idea of distributed cognition, where he argued that a naval crew's ability to successfully navigate was primarily due to environmental scaffolds that created 'observable representations' of cognitive systems. By distributing (and scaffolding) cognition, the crew is able to 'step inside' and manipulate traditionally internal cognitive representations. Thus, the cognitive demands of navigation are greatly reduced. It could even be said that culture is a type of cognitive scaffolding. Hollan, Hutchins and Kirsh (2000) argue that culture is an external structure that "shapes the cognitive processes of systems that transcend the boundaries of individuals" (p. 178).

So, simply put, cognitive scaffolding refers to external structures that change the cognitive demands of tasks. These external structures improve our cognitive powers—which would be significantly diminished in their absence—and contribute to our success in the world. Cognitive scaffolds come in many diverse forms and the ways we use them are widespread. We have also greatly benefited from downstream niche construction, passing down, inheriting and improving upon our scaffolds across generations (Sterelny, 2010). This is part of the reason Clark (1997) argues that we are set apart from other animals—at least cognitively—by our advanced ability "to create and maintain a variety of special external structures (symbolic and social-institutional)" (p. 179). However, perhaps due to the large quantity of scaffolding that improves our cognitive processes, there exists a tendency to view scaffolding as acting on agents in (primarily) positive and benign ways. I wish to argue that the opposite is also possible! In other words, there exist external structures that change the cognitive demands of tasks in ways that undermine the interests of one agent, while benefitting those of another. It is this hostile cognitive scaffolding that warrants further analysis.

This dissertation is divided into 4 chapters. Chapter 1 lays out some of the key groundwork for developing my later arguments. I begin by giving a brief historical overview of the term 'scaffolding'—discussing the word's transition from Soviet psychology to the cognitive sciences (§1.2) (Shvarts & Bakker, 2019). I then lay the philosophical groundwork for chapters 2 and 3. This includes characterising the features of ('benign') cognitive scaffolding—drawing primarily from Hutchins (1995a), Clark (1997) and Kim Sterelny (2010)—as well as providing some additional examples (§1.3). I use the term cognitive in an inclusive way so examples of affective scaffolding will also be discussed (Colombetti & Krueger, 2015; Piredda, 2019). I then further develop my characterisation of scaffolding by making a distinction between 'deep' and 'shallow' cases (§1.4). Shallow scaffolding primarily involves superficial cues, whereas deep scaffolding involves the scaffolding facilitating significant portions of the processing work. Next, I show that this deep scaffolding is possible by appealing to John Sutton (2010) and Richard Menary's (2006) complementarity principle (§1.5). According to the complementarity principle, when undertaking certain cognitive tasks, internal processes and external structures should be considered parts of a unified system. This explains deep scaffolding as extensions of our cognitive processes. In §1.6, Sterelny's (2010) three scaffolding dimensions (trust, entrenchment and sharing) will then be reviewed—explaining how we establish trustworthy and entrenched couplings with various scaffolds. I close the chapter by noting the abundance of benign examples in the literature and suggesting that hostile scaffolding should also be considered (§1.7).

In Chapter 2, I characterise hostile scaffolding by expanding on what is meant by the term 'hostility'—drawing from Sterelny (2003) (§2.2). Next, I elaborate on my usage of the term 'interests' (§2.3). These are relative to the agent performing a cognitive task and can be served or undermined. We are able to identify hostile scaffolding cases by asking "Who suffers?" (*Cui malo?)* and "Who benefits?" (*Cui bono?*) and identifying a beneficiary who is causally 'in the loop'.

The chapter then reiterates that hostile scaffolding is a genuine concern due to our extensive use of external structures for cognitive tasks (§2.4). Despite this, I describe (in more detail) how the potential negative effects of scaffolding are overlooked or underemphasised. The final aim of the chapter is then to review other relevant works that examine (cognitively) harmful external structures (§2.5). This includes the work of Jesper Aagaard (2020), Jan Slaby (2016), and Shen-yi Liao and Bryce Huebner (2020). Their work is highly insightful yet does not sketch the picture of hostile scaffolding (or deep scaffolding) as I do here. I suggest that my thesis develops useful terminology that can be applied to these projects.

Chapter 3 is the most crucial and demonstrates the existence of hostile scaffolding by providing everyday examples. I use the criteria developed in Chapter 2 to discuss sunglasses (Viola, 2022) (§3.2) and casino interior design and ambience (Friedman, 2000; Schüll, 2012) (§3.3.1) as examples of shallow hostile scaffolding. I then show that deep hostile cases are not just theoretical by analysing gambling machines that use player tracking systems and virtual reel mapping (Schüll, 2012) (§3.3.2) as well as Twitter's use of gamification techniques (Nguyen, 2021) (§3.4).

Finally, before concluding, Chapter 4 addresses future hostile scaffolding projects (§4.1). These include the ethical implication of deep hostile scaffolding, scaffolding's role in forming addictive behavior, other instance of gamification, 'racist scaffolding', and hostile scaffolding in developing technologies (such as VR, dark patterns and AI). I then conclude by reviewing the main arguments presented throughout the dissertation (§4.2).

# Chapter 1: What is Cognitive Scaffolding?

## 1.1. Introduction

The main aim of this dissertation is to make the case for a neglected type of cognitive scaffolding: hostile scaffolding. However, before we start that task, it is useful to establish a clearer understanding of scaffolding more broadly and to lay the groundwork for my arguments in chapters 2 and 3. As noted in the previous introduction, cognitive scaffolding refers to external structures that change the cognitive demands of tasks. By the end of this chapter, I will have expanded on this characterisation and set us up for an exploration of 'hostility' in Chapter 2. §1.2 begins by looking at the early development of 'scaffolding' in Soviet psychology and neurophysiology. By drawing from Anna Shvarts and Arthur Bakker (2019), I offer an account of how scaffolding came to be used in anglophone psychology and the cognitive sciences. Next, §1.3 gives an overview of scaffolding's presence in contemporary philosophy and cognitive science—reviewing the work of Edwin Hutchins (1995a), Andy Clark (1997) and Kim Sterelny (2003, 2010). I also note that my use of the word 'cognitive' includes affectivity and list some examples of affective scaffolding as well (Colombetti & Krueger, 2015; Piredda, 2019). In §1.4, I draw a distinction between 'shallow' cognitive scaffolding and 'deep' cognitive scaffolding. Shallow scaffolding primarily involves cue-based structures, whereas deep scaffolding facilitates significant portions of an agent's cognitive processing. This distinction is sometimes a matter of degree and need not be all or nothing. Mixed and borderline cases are possible and certain cases can be difficult to classify. The distinction is, nonetheless, useful for developing our understanding of cognitive scaffolding. Next, in §1.5, I ask how deep cognitive scaffolding is possible. I first frame scaffolding in the extended cognition literature, arguing that deep scaffolding is a type of cognitive extension, before reviewing the parity principle and extended functionalism (Clark & Chalmers, 1998; Wheeler, 2010). I conclude that these views are insufficient, since they do not account for the differences between internal and external vehicles. That said, deep scaffolding (as an extended cognitive processes) is still possible when we appeal to the complementarity principle—the idea that the internal processes and external structures used in cognitive processes should be considered parts of a unified system (Menary, 2006; Sutton, 2010). I then survey Sterelny's (2010) three dimensions (trust, entrenchment and sharing) and show how they offer additional explanations for how scaffolding establishes complementarity with our cognitive systems—by becoming trustworthy and entrenched (§1.6).

Lastly, I close the chapter by discussing the overabundance of benign examples in the scaffolding literature. I suggest that such an outlook is narrow sighted and briefly gesture towards the possibility of hostile scaffolding.

## 1.2. The Development of 'Scaffolding': From Soviet Psychology to Cognitive Science

Let us begin by looking at how the term 'scaffolding' developed, moving from its original use in Soviet psychology and neurophysiology to anglophone psychology and (later) cognitive science. Conventionally, scaffolding is defined as "poles and boards made into a temporary framework that is used by workers when they are painting, repairing, or building high parts of a building" (Collins, n.d.). Indeed, the everyday sense of the word refers to a (usually) temporary, physical structure that facilitates a task. However, in the 1970s, the English term also developed into a metaphor to describe psychological supports (Shvarts & Bakker, 2019). Shvarts and Bakker (2019) have done extensive research into this development and "address competing accounts about the origin of the term" (p. 4). According to them, the practice of talking about scaffolding for psychological purposes rather than physical tasks can be traced back to the 20th century Soviet psychologists Lev Vygotsky and Alexander Luria, and the Soviet neurophysiologist Nikolai Bernstein (Shvarts & Bakker, 2019). This would later develop into its more contemporary psychological usage (in the English language) by the psychologists David Wood, Jerome Bruner, and Gail Ross (1976). This section briefly outlines these developments and concludes by discussing why the term's usage in cognitive science is appropriately broad.

### *Scaffolding in Soviet Psychology & Neurophysiology*

There is a tendency among scholars to only reference Vygotsky as having first used the concept of scaffolding in a psychological context (Clark, 1997, 2003; Stephan, 2018). This supposed Vygotskian usage is in reference to the *zone of proximal development* (ZPD), which describes the initial stages of learning whereby a learner can only perform a task successfully with a supportive guide—usually a teacher (McLeod, 2019). The end goal is the development of independent, competent skills where the scaffolding is no longer required. However, according to Shvarts and Bakker (2019), aside from a single mention in his *Notebooks*, Vygotsky did not actually use the word 'scaffolding' in his writings (p. 6). Rather, due to similar principles at play, other authors have used the term "synonymously with the *zone of proximal development*" (Roth & Jornet, 2017, p.

122, as cited in Shvarts & Bakker, 2019). Vygotsky's single use of the term was during a time where he "was still searching for the correct name for the approach that he intended to elaborate" (Shvarts & Bakker, 2019, p. 7).

Vygotsky's contemporaries, Luria and Bernstein, would greatly contribute to developing the idea of scaffolding. The three men worked together from 1925 to 1927—with Vygotsky and Luria going on to collaborate (Shvarts & Bakker, 2019). It is understood that Luria wrote Chapter 3 of *Ape, Primitive Man, and Child,* which uses scaffolding to discuss ontogeny (Shvarts & Bakker, 2019, p. 8). The chapter describes how external supports—such as furniture or an adult's hand—are often used when children learn to walk. A child is "still surrounded, as it were, by the scaffolding of those external tools" because their abilities are not yet developed (Luria & Vygotsky, 1930/1992, p. 145, as cited in Shvarts & Bakker, 2019, p. 8). Similarly, Bernstein played a role in developing the concept through his theories of motor development. He suggested that agents sometimes learn by reducing their degrees of freedom (Shvarts & Bakker, 2019). A pianist, for example, may impose restrictions on their movements until greater competence is developed. The learner "rigidly, spastically fixed and holds the limb involved, or even his whole body, in such a way as to reduce the number of kinematic degrees of freedom which he is required to control" (Bernstein, 1940/1967, p. 108, as cited in Shvarts & Bakker, 2019, p. 9). When Bernstein later expanded on this idea, he noted that the newly learnt movements have to be executed "almost entirely under the control and correction of the leading level" and he likened these "surrogate corrections" to "temporary wooden scaffolds by means of which the future stone construction will be built" (Shvarts & Bakker, 2019, p. 9). The 'scaffolding', in this case, is the application of (higher level) corrections that, once mastered, operate in the background.

### *Scaffolding in Anglophone Psychology*

Shvarts and Bakker (2019) suggest that it was Bernstein's theory (with influences from Luria and Vygotsky) that would influence scaffolding's move to anglophone psychology. Wood, Bruner, and Ross—drawing from Bernstein—would notably bring the term to Harvard and Oxford in the 1970s (Shvarts & Bakker, 2019, p. 8). For them, 'scaffolding' referred to "the situation of teaching and learning" and was mainly an inter-subjective process involving another agent (Shvarts & Bakker, 2019, p. 10). Despite also expanding the notion of 'reducing degrees of freedom', the core properties of Bernstein's scaffolding were kept: (1) a temporary support that aided in the formation of new competencies; and (2) the process (typically) involved the interactions between a teacher and student, or the interplay of leading and background levels

(Wood et al., 1976). However, while Bernstein was more concerned with intra-subjective scaffolding (occurring in self-corrections), Wood et al. were interested in inter-subjective scaffolding (occurring in teaching and education) (Shvarts & Bakker, 2019). In this way, the use of scaffolding in anglophone psychology became more Vygotskian and analogous to ZPD, despite Wood et al.'s Bernsteinian influences.

### *Scaffolding's Move to the Cognitive Sciences*

Shvarts & Bakker (2019) note that, as scaffolding's usage has increased over the decades, "its meaning is becoming vague and polysemic" (p. 5). They argue that while it should be allowed to expand into other fields, the core idea of "a temporary adaptive support that forms a functional system with the learner" should remain (Shvarts & Bakker, 2019, p. 16). Contemporary cognitive science is not so restrictive. If scaffolding only referred to temporary supports, then it would be overly limiting and discount certain cases. When using a calendar, for example, our goal is not to reach higher levels of planning where we can (eventually) plan our days without it. Rather, by using the calendar, we are continually scaffolding our cognitive processes to aid in organisational tasks.

So, the field of cognitive science takes the idea of supportive external structures and applies it to a wider range of cognitive functions. The term's usage in this context is therefore appropriately more elastic and inclusive—not strictly concerned with temporary structures that are expected to be internalised. The scaffolding can be used indefinitely or as necessary parts of a task's performance. Returning to our characterisation: in the cognitive sciences, scaffolding refers to external structures that change the cognitive demands of a task. They can be temporary but that is not a necessary condition. This is the type of scaffolding I address throughout this dissertation. Hereinafter, when I refer to 'scaffolding', I am specifically referring to 'cognitive scaffolding'.

## 1.3. Scaffolding in Philosophy & Cognitive Science

We can now turn our attention to how scaffolding developed in modern cognitive science and philosophy. As noted previously, the literature around cognitive scaffolding has become increasingly vast and developed—particularly in the area of 4E (or situated) cognition[1]

[1] 4E, or situated, cognition refers to cognitive processes that operate outside of standard ('inside the head') conceptions of cognition and includes embodied, embedded, extended, and enacted cognition (Newen et al., 2018, pp. 3-4).

(Kiverstein, 2018; Menary, 2018; Newen et al., 2018). Inputting the words "cognitive scaffolding" into Google Ngram Viewer reveals a massive spike in its usage—an increase of almost eight-fold from 1994 to 2017. This section attempts to navigate the topic's ever-increasing body of literature, beginning with Hutchins' (1995a) idea of distributed cognition. I then discuss Clark (1997) and Sterelny (2003, 2010)—who were influenced by Hutchins (and others[2]) and raised the profile of cognitive scaffolding in the philosophical discourse. I close this section by showing that, since I use 'cognitive' in an inclusive way, affective scaffolds should be included in our scaffolding discussion (Colombetti & Krueger, 2015; Piredda, 2019).

### *Hutchins & Distributed Cognition*

Discussions around distributed cognition and scaffolding greatly overlap—since they both involve external structures that transform cognitive processes. Hutchins (1995a) first developed the theory of distributed cognition and he would greatly shape the scaffolding literature that followed (Clark, 1997). Jesper Aagaard (2020) notes that "distributed cognition is not an empirical type of cognition; it is a theoretical perspective on all forms of cognition" (p. 5). Hutchins' (1995a) book *Cognition in the Wild* mainly focuses on navigation aboard naval vessels. There, Hutchins extended our cognitive models to not only include agents' internal cognition but also the "effort of the technologically equipped and socially organized crew" (Walter, 2018, p. 326). The core idea was that naval navigation can be characterised as involving 'observable representations' of cognitive systems. As Hutchins (1995a) describes:

> On this view of cognitive systems, communication among the actors is seen as a process internal to the cognitive system. Computational media, such as diagrams and charts, are seen as representations internal to the system, and the computations carried out upon them are more processes internal to the system. Because the cognitive activity is distributed across a social network, many of these internal processes and internal communications are directly observable… With systems of socially distributed cognition we can step inside the cognitive system, and while some underlying processes (inside people's heads) remain obscured, a great deal of the internal organization and operation of the system is directly observable. (pp. 128-129)

Ship navigation systems therefore distribute cognitive processes across the crew and external structures (tools such as diagrams and charts) in order to successful complete tasks—with internal and external processes working together. By being able to 'step inside' this system,

[2] This includes thinkers such as David Kirsh and Paul Maglio (1994), to name a few.

certain (external) cognitive processes become observable and manipulable. We cannot attribute successful ship navigation of this kind to individuals alone; we must also consider the manipulation of external structures and the organisational systems in place (Hutchins, 1995a, p. 173). As already noted, the idea of cognitive distribution greatly overlaps with scaffolding.[3] In some scaffolding cases, we see a cognitive loop forming: cognition begins in the head and moves outside, where it is transformed by external media, before looping back inside the agent. For example, when solving a distance-rate-time problem, navigation practitioners can make use of a three-scale nomogram—a diagram that shows the relations between a number of variable values—to calculate speed (Hutchins, 1995a, p. 150). The task solver finds the time and distance values on the nomogram and marks them. When drawing a straight line through the marked lines, it intersects with a third value—indicating the speed. This external structure thereby transforms and simplifies calculations.

The environment of a naval crew is not only scaffolded with navigational artefacts, such as diagrams and charts, but also social dimensions and cultural learning. Hollan, Hutchins and Kirsh (2000) state that "the study of cognition is not separable from the study of culture, because agents live in complex cultural environments" (p. 178). A novice navigator sometimes learns by observing a "set of culturally prescribed behaviors" and there are often expectations "about who needs what kinds of help in learning the job" (Hutchins, 1995a, p. 280). Novices observe (and are told) what constitutes appropriate action, as well as which tasks are too difficult for them to perform. More broadly, Hollan et al. (2000) note that "culture provides us with intellectual tools that enable us to accomplish things that we could not do without them" (p. 178). Scaffolding can therefore comprise of physical tools as well as social and institutional dimensions. Even though Hutchins does not always refer to external structures as 'scaffolding', they can often be characterised as reducing the cognitive demands of tasks.

### *Andy Clark's Account of Scaffolding*

Clark (1997) draws heavily from Hutchins (1995a) and is one of the earliest philosophers to bring the philosophical discussion of scaffolding, and later extended cognition (Clark & Chalmers, 1998), into greater prominence. I review Clark's (1997) notion of scaffolding, as introduced in *Being There*, and discuss how it is in line with my characterisation (while also

[3] Hutchins (1995a) also describes how various systems "scaffold" (in a Vygotskian sense) the learning process of practitioners (p. 280).

expanding it). I also discuss David Kirsh and Paul Maglio's (1994) work on epistemic actions, as they, too, greatly influenced Clark (1997).

*Being There* explores the dynamic relationship between mind, body and world (Clark, 1997). When it comes to performing certain cognitive processes, the three are often so interconnected that drawing distinct boundaries between them becomes difficult. Scaffolding, for Clark (1997), plays an important role in blurring these boundaries[4] and refers specifically to our "exploitation of external structure" (p. 45). His initial examples include how the "structure of a cooking environment (grouping spices, oils, etc.)" can act "as an external memory aid" (Cole et al., 1978, as cited in Clark, 1997, p. 46), and how a child uses "special eating utensils that reduce [their] freedom to spill and spear" (Valsiner, 1987, as cited in Clark, 1997, p. 46). We can start expanding on my characterisation by considering how scaffolding can be involved in epistemic actions. Kirsh and Maglio (1994) describe epistemic actions as actions that change an agent's computational state by making "mental computation easier, faster, or more reliable" (pp. 513-514). In other words, an epistemic action is one "whose primary purpose is to alter the nature of our own mental tasks" (Clark, 1997, p. 64). While our brains perform some cognitive processes, others are delegated to manipulations of external structures. For example, Kirsh and Maglio (1994) detail how computer aided block rotation in the videogame *Tetris* make perceptual and cognitive problems more easily and reliably solvable. A computer with a rotation button allows us to fit geometric shapes (called 'zoids') into the appropriate spaces roughly three times faster than 'inside the skin' methods (Kirsh & Maglio, 1994, pp. 527-533).[5] By manipulating these structures to perform certain tasks, we are performing epistemic actions. Other epistemic actions described by Kirsh and Maglio (1994) include:

> Memory-saving actions such as reminding, for example, placing a key in a shoe, or tying a string around a finger; time-saving actions such as preparing the workplace, for example, partially sorting nuts and bolts before beginning an assembly task in order to reduce later search. (p. 515)

Our scaffolding characterisation can therefore include how external structures "alter the nature of our own mental tasks" by sometimes allowing us to perform epistemic actions (Clark, 1997, p. 64).

[4] Scaffolding specifically blurs the boundaries between mind and world.
[5] This is roughly 300ms, when aided by a computer, compared to 1000ms, when unaided (Kirsh & Maglio, 1994, pp. 527-533).

Let us review memory-saving actions in more detail. We often leave objects in specific locations to act as reminders. Suppose you always place a rubbish bag near your front door the night before garbage day, or you leave an empty bottle of olive oil on the kitchen counter when you need to buy another. By making use of intelligent space management, you reduce memory demands by prompting additional actions. Hollan et al. (2000) note:

> Space is a resource that must be managed, much like time, memory, and energy. Accordingly we predicted that when space is used well it reduces the time and memory demands of our tasks, and increases the reliability of execution and the number of jobs we can handle at once (p. 190).

One may be able to remember to do certain tasks without these scaffolds but doing so is likely a less reliable and a more cognitively demanding exercise. Furthermore, the ubiquity of scaffolding in our daily lives cannot be understated. Agents "constantly create external scaffolding to simplify [their] cognitive tasks", even though they may not be aware of extent to which they do so (Hollan et al., 2000, p. 192).[6]

Clark would likely agree with Hollan et al.'s (2000) assessment. Chapter 9 of *Being There* not only discusses how scaffolding is used when performing epistemic actions, but also its prevalence (Clark, 1997). External structures can distribute cognition across physical domains (such as when using pen and paper to performing mathematical calculations) or, as we discussed earlier, through the widescale implementation of social and institutional processes. As Clark (1997) puts it: "We use our intelligence to structure our environment so that we can succeed with less intelligence. Our brains make the world smart so that we can be dumb in peace" (p. 180). Clark (1997) asserts that we are not 'smarter' than the generations that came before; rather, scaffolding allows us to better cope with the burden of increasingly complex cognitive tasks. Clark's (1997) view, much like Hutchins (1995a), also demonstrates that scaffolding makes collective cognitive accomplishments possible. When operating with scaffolding in place, we are able to: (1) perform advanced cognitive (and coordinated) tasks, all while "reduc[ing] the loads on individual brains"; and (2) spread collectively achieved wisdom and knowledge among individuals (Clark, 1997, p. 180). Scaffolding therefore forms a type of 'dissipated reasoning'. This is in line with the previously discussed idea that "culture shapes the cognitive processes of systems that transcend the boundaries of individuals" (Hollan et al., 2000, p. 178). Culture is an

[6] I return to this point in Chapter 2.

external structure that has already created partial solutions to problems, ensuring that new generations do not have to start from scratch.

As another example, consider, as Clark (1997) does, the organisation of firms and corporate office structures. Here, we not only see the more typical manipulations of external structures (such as the use emails or slips of paper), but socio-institutional practices as well. Clark (1997) lays out this process as follows:

> Daily problem solving, in these arenas, often involves locally effective pattern-recognition strategies which are invoked as a result of some externally originating prompt (such as a green slip in the "in" tray, discharged in a present manner) and which leave their marks as further traces (slips of paper, e-mail messages, whatever) which then are available for future manipulations within the overarching machinery of the firm. (pp. 185-186)

By using norms, policies and practices, organizational structures can support the completion of complex cognitive tasks, all while reducing the cognitive demands of the individuals involved. Tasks are completed and then further coordinated with other individuals and departments, allowing the whole organization to function effectively. As Hollan et al. (2000) note, "since social organization—plus the structure added by the context of activity—largely determines the way information flows through a group, social organization may itself be viewed as a form of cognitive architecture" (p. 177). As also seen with Hutchins' (1995a) work, cognition is thus distributed across the organization.

Clark's (1997) work in *Being There* helped raise the profile of scaffolding in philosophy. He demonstrated that, when considering our cognitive accomplishments, scaffolding (often) deserves a large portion of the epistemic credit.[7] Even though Clark (1997) does not give a fully detailed description of scaffolding, as I do in this dissertation, his analysis is in line with my characterisation. Clark (1997) also makes the link between scaffolding and epistemic actions clear, while emphasising, like Hollan et al. (2000), its widespread use in many of our daily tasks.

### *Kim Sterelny's Account of Scaffolding*

Sterelny is another notable figure in the philosophical discussion around scaffolding. One of the main aims of his paper, *Minds: Extended or Scaffolded?*, was to show why the scaffolded mind is more interesting than the most compelling and plausible cases of extended mind (Sterelny,

[7] We also see similar ideas developed in Clark and David Chalmers' (1998) later views on the extended mind: some cognitive processes are not only 'in the skull'; they can extend to external structures.

2010).[8] But, setting that matter aside, we can use this paper to expand on our understanding of scaffolding. Part of Sterelny's (2010) answer to how agents come to act effectively in varied and changing environments is through the process of niche construction. Even though species do adapt to their environments over time, they can also adapt their environments to them. One example can be found in how beavers build dams to influence the flow of running water (Sterelny, 2010, p. 470). Niche construction can also take the form of epistemic actions, since "agents alter the informational character of their environment in ways that make crucial features more salient" (Sterelny, 2010, p. 470).[9] Additionally, some animals, like beavers, develop a form of "downstream niche construction" (Sterelny, 2010, p. 470). Their constructions are not just useful to them, but other generations as well. As previously discussed, humans are a notable example of this and we also possess profound social learning abilities. By using our capacity to learn and pass down constructed niches, we transmit "ecological and technical expertise" across generations (Sterelny, 2010, p. 470). This is can be seen in the panoply of external structures we pass on[10], as well as in the ways we learn to construct and use tools (in apprentice systems, for example) (Sterelny, 2010). All of the above cases count as scaffolding and Sterelny (2010) lays out the scaffolded mind hypothesis as follows:

> The scaffolded mind hypothesis proposes that human cognitive capacities both depend on and have been transformed by environmental resources. Often these resources have been preserved, built or modified precisely because they enhance cognitive capacity. (p. 472)

We saw earlier that Clark (1997) also highlighted the importance of scaffolding in our everyday lives. Sterelny (2010) links certain scaffolds to the process of niche construction and further emphasises the idea that our cognitive capacities rely on and have been dramatically changed by our use and inheritance of scaffolding.

Sterelny's (2003) other work offers additional insights into the evolutionary role of scaffolding. He presents many accounts of how animals interact with and identify features of their environment in order to select appropriate actions (Sterelny, 2003). The emergence of detection systems is one of the most important evolutionary developments on this front. In their simplest form, these are "single-cued discriminatory mechanisms" which allow an organism with less robust cognitive architecture to track and respond to certain features of the environment

[8] This is because the scope of scaffolding is wider than that of the extended mind project (Sterelny, 2010).
[9] We saw this is in previous examples—like the cook who lays out ingredients to act as memory aids.
[10] Road markings and signs, for example, are likely to be used by the next generation.

(Sterelny, 2003, p. 14). Because some animals may not have fully adequate internal models of the world, other mechanisms had to develop in order to cope with detection. Without complex individual intelligence or representations, scaffolding (in some cases) became a requirement in order to identify environmental features and successfully perform certain actions. Hawks are one such animal that scaffold their cognitive processes. For example, a hawk "chooses a roost which maximises its view of its hunting territory", thus allowing for easier prey detection (Sterelny, 2010, p. 470). By choosing a strategic position for its nest, the hawk reduces the cognitive complexity of food detection. Ants, too, scaffold their environments by marking out their world with pheromone trails. Sterelny (2003) draws from Hölldobler and Wilson (1990) to lay out the following:

> Detection systems can also generate adaptive behaviour when agents 'store information in the world' rather than in their brain. Adaptive behaviour is scaffolded by agents physically engineering their environment. They act on their environment so that it subsequently generates cues that support adaptive responses. Ant pheromone trails fit this picture of organisms storing information in the world. Such external storage can be very rich in information. Those trails carry information about direction and distance, and the number of ants using it carries information about the value of the food resource. (p. 19).

The pheromone trails scaffold the ants' environment by allowing them to more accurately discern environmental information, respond more appropriately to cues, as well as co-ordinate and act collectively as a group. In this way, an ant's brain is able to perform tasks that surpass their 'inside the skull' cognitive abilities. To rephrase Clark (1997), they make the world smart so they can be dumb in peace (p. 180). Even though Clark (1997) argues that our more advanced and prolific ability "to create and maintain a variety of special external structures" is what sets us apart from other animals (p. 179), which is largely unchallenged by other authors, Sterelny (2003, 2010) demonstrates that scaffolding can still play important roles in how other animals accomplish various tasks.

### *Affective Scaffolding*

Most of my discussion around distributed cognition and scaffolding has so far focused on traditional (narrowly) cognitive supports—i.e., perception, memory and inference. But I use the word 'cognitive' to include affective processes as well. Various scholars have argued that the scope of scaffolding should be expanded to include affectivity (Colombetti & Krueger, 2015;

Griffiths & Scarantino, 2009; Piredda, 2019). Giulia Piredda (2019) notes that, in recent times, "not only has the nature of affective states come under scrutiny, but also the interaction between our affective life and the environment we inhabit" (p. 2). This has prompted arguments for extending the 'situated' label to included affectivity—where affective states are also supported by 'outside the head' structures. In a similar vein to what Hutchins (1995a) proposed with distributed cognition, Griffiths and Scarantino (2009) present a theory of situated affectivity. Here, they argue that emotions should not be seen as "internal states that provide an organism's decision making system with information about a certain situation", but as "complex events constituted by a dynamical unfolding, developed in deep interaction with the physical and the social environment" (Piredda, 2019, p. 2). In this way, the situated affectivity picture can show us the extent to which our lives are also affectively scaffolded. This section will briefly present some examples of affective scaffolds. As we consider each case, we will see that it is appropriate to extend the scaffolding label to them as well. Colombetti and Krueger (2015) argue, convincingly, for this inclusion by applying Sterelny's (2010) previously discussed model to the creation of affective niches—i.e., the combination of scaffolding with the situated view of emotions. But for this section, I instead draw mainly on Piredda (2019). She acknowledges Colombetti and Krueger's (2015) views while also expanding them.

Piredda (2019) states that affective artefacts "alter the affective condition of an agent" and classifies to them as a subclass of affective scaffolding (p. 2). Let us begin by considering comfort objects. A comfort object (or security blanket) is something we are all likely familiar with. In children, this usually takes the form of a physical object, such as a teddy bear or toy, and helps agents access calm mental states. Linus, from the *Peanuts* comic strip, is generally depicted holding a blanket, which keeps him calm and assured. As Piredda (2019) notes, "the presence or absence of the blanket affects Linus' affective condition, contributing to its regulation" (p. 3). For adults, comfort objects can take many forms as well. Jewellery (such as a wedding ring) often regulates our affective states by reminding us of connections to family and loved ones (Piredda, 2019); and photos of home, family or friends can reassure or offer us comfort in foreign environments. Additionally, even though a handbag has the practical use of storing items—which can make one feel prepared and confident—its presence alone seems to offer wearers a sense of comfort and security. Colombetti & Krueger (2015) argue that a handbag offers a sense of "confidence, power, and security" as well as "corresponds to, and completes, a certain self-styled body image" (pp. 1163, 1165). Leaving the house without one's handbag may elicit feelings of unease and distress or the sense that one is "not complete" (Kaufman, 2011, p. 157, as cited in Colombetti & Krueger, 2015, p. 1165). Similarly, clothes (such as a suave suit) scaffold our affective states by making us feel more secure, confident or attractive. An inappropriate fashion decision can, alternatively, have the opposite effect and lead to us feeling self-conscious. Finally, music, although not an artefact, also scaffolds affectivity. A fitting workout playlist can energize us while running or in the gym, an upbeat melody may elicit feelings of happiness in times of depression, and classical music may allow us to focus more attentively during long study sessions. Conversely, an ill-fitting song may hinder these activities instead of aid them.

What is clear from Piredda (2019) (and others[11]) is that affective scaffolds play important roles in changing the cognitive demands of our lives. As Piredda (2019) notes, "we all know from personal experience how valuable such objects can be to us, especially in particular moments of our lives when our identities are undergoing a process of construction" (p. 3). The use of the word "construction" is particularly apt and is indicative of the role (and importance) of affective scaffolding in our emotional lives and self-development. Additionally, the management of affective states may also influence other, more traditional cognitive processes. Music that allows for calmer mental states may lead to better writing, for example, or influence decision making (which will be discussed in §3.3.1). I return to affective cases in my later discussions in Chapter 3.

## 1.4. Deep & Shallow Scaffolding

At this point, we can draw a distinction between two types of scaffolding: 'deep scaffolding' and 'shallow scaffolding'. Let us begin with the latter. We can observe that many of our previous examples involve agents using cues. The success of a roosting hawk in spotting its prey, or in ants identifying resources with pheromones, is largely due to the animals picking up on and acting appropriately to cues. Similarly, in human examples, object reminders, road markings and public signs involve following and acting on cues. In all of these examples, most of the actual processing work is still being done by the agent. Cue-based scaffolding is therefore 'shallow' with regards to the cognitive role it performs. However, there may be 'deep' cases where a significant portion of an agent's processing work is facilitated by the scaffolding itself. Consider how we use a calendar or daily planner. Not only does this aid in our ability to remember certain events and plans, but it also offloads substantial cognitive processing to the scaffolding. It structures and lays out the days and weeks of the month in ways that allow you to more effectively visualise and access your time. Imagine a colleague asks whether you are free for a meeting on Thursday, one

[11] This includes Colombetti and Krueger (2015), and Griffiths and Scarantino (2009).

month from now. Supposing you have a demanding schedule, you would most likely be unable give immediate answer with your 'on-board' cognitive resources alone. But, if you were diligent with updating your calendar, you could glance over to the suggested date and give an almost immediate answer. If the answer is yes, you can then log that event—which aids in remembering or avoiding conflicting plans. Compare this to the (unrecommended) process of jotting down each day of your upcoming month on separate pieces of paper and placing them in a bag marked November. After rummaging through your bag, you may eventually be able to answer your colleague. It would, however, take considerably longer. Using the calendar is far more effective because the calendar's layout is that of a computational[12] structure that preserves certain relationships. Being adjacent in space corresponds to being adjacent in time and regular columns can reinforce the cycle of weekdays and weekends. This provides a structured, visual relationship of your time and schedule and allows you to perform cognitive tasks that would otherwise be more difficult.

Next, consider how we perform mathematical calculations with scaffolding. An often-used example is the use of pen and paper in solving long multiplication or division problems. The pen and paper allow you to transfer and store numerical information while also simplifying a problem. We can then use mathematical norms and practices to calculate an answer. As Clark (1997) explains:

> What we achieve, using pen and paper, is a reduction of the complex problem to a sequence of simpler problems beginning with 2 × 2. We use the external medium (paper) to store the results of these simple problems, and by an interrelated series of simple pattern completions coupled with external storage we finally arrive at a solution. (p. 61)

While the use of pen and paper is a good instance of scaffolding facilitating significant computational/ processing work, I will focus on a similar (although outdated) example: a slide rule. Slide rules consist of three parts: a body, slide, and cursor. They typically perform division, multiplication, square and square root functions—although they can perform other tasks with different methods or modifications. Let us perform a hypothetical slide rule calculation. Suppose you wish to calculate 9645 ÷ 650; you start by moving your attention to the far-right side of the slide rule and placing the cursor at approximately the value of 9645. You then move the middle slide to a new position so the value of 650 and the cursor intersect. The left end of the slide rule then gives you the approximate answer of 148. The final step is the only 'inside the head'

[12] Some may argue that this is not 'computational' per se (Piccinini & Scarantino, 2010), but I use the term in a very loose and inclusive way when describing certain cognitive processes.

calculation and that is determining that the decimal point is after the 14. You are then left with the answer of 14.8. For their time, slide rules were remarkably adaptable and could perform mathematical computations all while allowing users to visualize the computations being performed. The slide rule and calendar both enable the offloading of significant cognitive processing. This is what is meant by 'deep': the scaffolding itself facilitates (or offloads) large portions of the cognitive work, which is external to the agent.

So, by looking closely at the above examples, we can mark a distinction between two types of scaffolding. The first I will refer to as 'shallow scaffolding' and the other as 'deep scaffolding'. As explored earlier, shallow scaffolding can be typified in the examples of the hawk roosts, ant pheromone trails and object reminders. These involve cue-based external structures, which then prompt additional cognitive functions. Once the hawk detects prey from its roost, it can engage in the act of catching it. Likewise, pheromone trails lead ants to identify something as food. But, once those processes are completed, the hawk's and ants' other cognitive faculties take over and act appropriately. Similarly, our object reminders also cue (remind) us to perform further tasks. However, this only acts as a memory prompt and the fulfilment of the task—taking out the trash or buying another bottle—is completed with other cognitive systems. In all of these cases, the scaffolding begins and ends with the cue. Deep scaffolding, on the other hand, is where the scaffolding facilitates significant 'outside the skull' processing. The scaffolding's structure preserves certain relationships and allows us to successfully perform cognitive tasks—provided we use the appropriate norms and practices. The calendar and slide rule fit this framework well. The structure of a calendar allows you to schedule your time manners that are far easier to read than the alternatives; and a slide rule enables you to complete certain mathematical calculations. When events are slotted into a calendar or when we move the cursor and slide of a slide rule, we perform new cognitive tasks (processing/ calculation) in collaboration with the artefacts. Hutchins' (1995a) example of the three-scale nomogram (as discussed in §1.3) also fits here, as it facilitates and simplifies external calculations due to its structure. These deep cases all involve a cognitive loop: cognition begins in the head and then moves outside, where it is significantly transformed by the external structure, and then loops back inside the brain. Furthermore, our current examples of deep scaffolding seem to be exclusive to human cognition. A single ant using a pheromone trail is an instance of shallow scaffolding, but a whole colony of ants coordinating their behaviour could be deeper. But this does not mean that possible 'deep' ant scaffolding operates in the same ways as deep human cases, just that there could be more external processing work being done.[13] It remains unclear if other animals are capable of engaging with deep scaffolding in the same ways as we do. But the potentially 'deep' ant example does reveal that the distinction between 'deep' and 'shallow' scaffolding is sometimes a distinction of degree. As such, the claim that some scaffolding is deep will sometimes be tricky to defend due to certain cases being harder to discern.

Finally, consider how some cases of affective scaffolding could also sit somewhere between the deep and shallow distinction. Linus' blanket may be a cue that results in a calm state of mind, but an emotional song or a well-constructed narrative could be 'deeper'. Imagine that you watch a movie that very effectively conveys an emotional message; for example, that the environment should be protected. This then resonates with you and triggers feelings of environmental protectiveness. You may feel a mix of emotions (love, anger or regret) and you decide to continually rewatch the film to renew these feelings and recall your responsibilities to the planet. This does not seem like a case of shallow scaffolding, as the movie is not a surface level cue. Instead, it offers a structured format whereby a series of experiences are organised in time to bring about certain affective responses. The order (in which events are structured) plays an important role in eliciting these emotional responses. Had *Hamlet* begun with our protagonist killing Claudius, it would have likely affected the audience differently. Similarly, the tempo of a song or the order and timing of a good joke's delivery can elicit different affective reactions. I am not claiming that these affective examples are performing cognitive processes/ calculations per se, as they are not equivalent to the calendar or slide rule examples, but simply that they are among the cases which are harder to pin down as deep or shallow.

## 1.5. Deep Scaffolding: How Is It Possible?

If scaffolding is deep, then some of the cognitive processing is happening 'outside the skull'. At this point, a number of questions arise: How is deep scaffolding possible and what theoretical framework can make sense of it? This section will attempt to answer these questions by appealing to the complementarity principle (or thesis) (Menary, 2006; Sutton, 2010). I begin this section by laying the groundwork for 4E cognition, specifically extended cognition, before explaining how this view is (often) defended by extended functionalism (Wheeler, 2010). I then discuss the parity principle as presented in Clark and Chalmers' (1998) *The Extended Mind*. We cannot, however, insist on explaining deep scaffolding through that version of the parity

[13] In other words, it may involve more external cognitive work than just reacting to simple cues.

principle alone, since external structures do not always work in the same ways that brains do. I then move to Menary (2006) and Sutton's (2010) work on complementary. Here, I use the complementarity principle to buttress the parity principle and argue that complementarity satisfactorily answers how deep scaffolding is possible.

### *4E Cognition & Functionalism*

In §1.3, I briefly discussed 4E (or situated) cognition. Before moving on to the parity principle and complementarity principle, a brief expansion on that field is required. The philosophical discussion around 4E cognition emerged in the 1990s and challenged the classical cognitive sciences—i.e., the representational and computational model of cognition (RCC) (Newen et al., 2018). To briefly recap, 4E cognition is the idea that cognition can also occur 'outside the head', which follows on from Hutchins' (1995a) picture of distributed cognition. 4E cognition theories argue that cognitive processes can be embodied, embedded, extended and enacted (Newen et al., 2018, pp. 3-4). As previously noted with the calendar and slide-rule examples, deep scaffolding forms a cognitive loop between the agent and external structure: we use the external structures to perform cognitive processes and the scaffolding facilitates significant portions of the processing work. This can be seen as an extended cognitive process so 'extended cognition' is the 'E' I will focus on.

The picture of extended cognition is (often) defended by appealing to functionalism (Wheeler, 2010; see also Clark, 2008). So, what is functionalism and why should we buy into it? In its traditional formulation, functionalism states that:

> If psychological phenomena are constituted by their causal-functional roles, then our terms for mental states, mental processes, and so on pick out equivalence classes of different material substrates, any one of which might in principle realize the type-identified state or process in question. (Wheeler, 2010, pp. 3-4)

According to functionalism, what makes something a mental state (such as pain or desire) does not depend on "its internal constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part" (Levin, 2018). Part of the reason functionalism is so appealing is that it allows for multiple realizability: "that a single mental kind (property, state, event) can be realized by many distinct physical kinds" (Bickle, 2020). For example, if we imagine encountering a Martian whose brain structure and make up is significantly different to ours (perhaps it is made of green gel), we seem to have the intuition that (so long as it has the same functional responses to systemic inputs and outputs) the Martian should be considered a genuine cognizer (Levin, 2018; Wheeler, 2010). This is known as the Martian intuition. As Wheeler (2010) notes, multiple realizability is not just expressed as a thought experiment for what might possible, but it seems to play a vital role in our own (non-Martian) world (p. 6). A variety of species seem to display convergent evolutionary traits which, despite having different biological substrates, perform the same function. Wheeler (2010) points to two distinct enzymes in vertebrates and fruit flies; both have the same function of breaking down alcohol, yet they have dramatically different chemical make ups (p. 6). He argues that if such 'fine grained' multiple realisability is possible here, then it should also be possible for certain mental states and cognitive processes to be realised through different material substrates[14] as well (Wheeler, 2010).

### *The Parity Principle & Extended Functionalism*

The parity principle is considered by Clark (2008) to be a "simple argumentative extension" of a "commonsense functionalism"[15] regarding mental states (p. 88) and was first introduced in Clark and Chalmers' (1998) highly influential paper *The Extended Mind*. By expanding on Hutchins' (1995a) ideas, as well as Kirsh and Maglio's (1994) work on epistemic actions, Clark and Chalmers (1998) applied the distributed cognition picture to minds.[16] The extended mind theorist argues that the external structures used to aid mental processes, such as belief forming, also deserved a spread of the epistemic credit. We should therefore extend the classification of our mental processes to include those external structures that are involved in certain tasks.[17] This placed a new focus on the situatedness of the mind rather than just cognition. But, perhaps most importantly, Clark and Chalmers (1998) introduced a rough philosophical framework for discussing cognitive extension that would be adopted, criticized and improved upon by themselves and their contemporaries (Clark, 2008; Menary, 2006; Sutton, 2010; Wheeler, 2010). The parity principle would become one of the paper's most discussed (and controversial) ideas. It states that: if we are performing a cognitive task using an external structure and, were this task done exclusively in the head, we would have no problem giving it cognitive status, then we should consider that external system cognitive as well (Clark & Chalmers, 1998). Or, put more simply, "the location per se of a process doesn't determine whether it counts as part of how the mind works" (Hurley, 2006, p. 20). Clark and Chalmers (1998) illustrate the parity principle with

[14] These substrates need not be limited to internal systems.
[15] "Commonsense functionalism" as described and endorsed by Braddon-Mitchell & Jackson (2007).
[16] Clark and Chalmers (1998) are not only concerned about "cognitive processes" but also other mental states. For example, how external couplings could be actively involved in belief forming processes.
[17] This is a type of *active* externalism: coupled systems where the environment plays an active role in cognitive processes (Clark, 2008; Clark & Chalmers, 1998). If we removed the external component, then our behavioural competence would decrease.

a thought experiment. They compare Inga (a fully functioning cognitive agent) with Otto (an Alzheimer's patient). To summarise, both Inga and Otto need to use a memory system to verify their belief that an art exhibit is taking place at the Museum of Modern Art (MoMA) on 53rd street. Inga uses her internal biological memory while Otto uses a diligently updated notebook. Clark and Chalmers (1998) argue that Inga's memory and Otto's notebook play (functionally) the same role in verifying their beliefs (pp. 12-14). Under the parity principle, Otto using his notebook to recall and verify that MoMA is on 53rd street, should be considered a genuinely cognitive process in the same way as Inga's internal memory. But we need not focus on the nuances of belief forming processes for our scaffolding purposes. What is most important to consider is the function of memory, which both Inga's biological brain and Otto's notebook serve, despite being realised through different substates. As scaffolding, Otto's notebook is a portable structure that changes (in this case fills in for) the cognitive demands of remembering by allowing Otto to record and access information. *The Extended Mind* concludes that, by following the parity principle, it may be possible for us to start thinking of minds as extending outside the skull and into the world (Clark & Chalmers, 1998). However, the idea that the mind extends, could be problematic. In an interview on *Closer to Truth*, Clark said:

> Certainly when I talk to certain cognitive scientists, the notion that cognition might extend is taken to be reasonably unproblematic. People start to worry when you say the mind extends, and I think that's because mind and consciousness are being kind of tied together. So, it might very well be that a reasonably deflationary understanding of the extended mind claim is warranted. (Kuhn [Closer to Truth], 2020)

Given my focus on scaffolding, a 'reasonably deflationary understanding of the extended mind' seems to be the more productive approach. I will not discuss mind extension, as it relates to consciousness or certain mental states (such as of desiring), but limit my discussion of deep scaffolding to cognitive (and affective) processes.

The version of functionalism that explicitly supports the parity principle is extended functionalism. The reason cognition can be 'extended' is because the chains of subsystems don't all have to be in the head to be considered cognitive, they can also occur outside of it by virtue of their functional properties. By endorsing the parity principle and extended functionalism, the ultimate goal for the extended cognition theorist is to allow for extended systems to have genuine cognitive status (in the strong sense)—i.e., "the constitutive dependence of mentality on external factors" (Wheeler, 2010, p. 2). Wheeler (2010) praises the parity principle and extended functionalism for not privileging the human inner as a benchmark for cognition—what he calls "neural or carbon chauvinism" (p. 3). Instead, they present independent standards for cognition which are "locationally uncommitted" (Wheeler, 2010, p. 11). That said, in parity principle and extended functionalist examples, the external structures must still be coupled with a cognitive agent (with a task) to be considered cognitive. We are not arguing that these processes can be cognitive without a coupling.[18] It is here that a problem arises. On their own, the parity principle and extended functionalism do not explain how the couplings between agents and external structures occur. Furthermore, internalist critics often argue that: (1) because internal and external processes are different, and are realized in different ways, they need to be treated as two different kinds; and (2) external vehicles (mediums) have too much variety in how they process information and, therefore, do not process information in genuinely cognitive ways[19] (Adams & Aizawa, 2001, 2006; Rupert, 2004). If these criticisms hold true, it would not only be damning for extended mind theorists but also my deep scaffolding arguments. If cognitive processes cannot extend outside the skull, or we cannot sufficiently explain agent-scaffolding couplings, then it is difficult to see how deep scaffolding could be considered genuinely cognitive parts of agents' processing.

### *Cognitive Integration through Complementarity*

So, how do we answer these criticisms without abandoning extended cognition? And, most crucially, how do we explain the couplings between agents and deep scaffolding? I now move on to the heart of answering how deep scaffolding is possible: John Sutton (2010) and Richard Menary's (2006) complementary principle. Here, we downplay the matter of parity and move closer towards connectionism. The complementarity principle argues that, because we sometimes have a strict complementarity between external structures and internal cognitive processes, they should be considered parts of a unified system. By appealing to an integrationist view of cognition, not only can we answer the internalist, by acknowledging the differences between internal and external vehicles, but we can also offer a solid explanatory picture for how these vehicles jointly perform cognitive processes.

Sutton (2010) argues strongly for the complementarity between certain external and internal processes and supports a hybrid theory of cognition. He lays out two main waves in the

[18] The agent must then also apply the appropriate norms and practices to use the scaffolding successfully.

[19] Rob Rupert (2004), for example, argues that certain cognitive functions, such as memory, have features that cannot be realized through external means. In the case of memory, Rupert (2004) considers the generation effect—where information is better remembered when generated by an agent rather than simply read—and how this does not occur in the same way as when one uses a notebook. This would constitute to a breakdown of the parity principle, since human memory functions cannot be performed 'outside the skin' (Rupert, 2004). But, as Wheeler (2010) points out, what Rupert (2004) is doing is taking a very specific human cognitive trait and making a generalisation about how all core cognitive traits should share these specific features.

extended cognition literature: wave 1 is based on the previously discussed parity principle, and wave 2 pursues the project of complementarity—which is a more well-developed and defendable position. Menary (2006), much like Sutton (2010), follows this second wave of extended cognition. For Menary (2006), the extended mind hypothesis (that Clark and Chalmers (1998) initiated) is part of a more radical project of 'cognitive integration'—where "internal and external vehicles are integrated into a whole" (p. 329). He also argues for complementarity over parity, since internalists have "attack[ed] a flawed comparative version" of the parity principle (Menary, 2005, p. 333). In this way, both Sutton (2010) and Menary (2006) support the complementarity principle as the stronger argument for extended cognition.

In his discussion of cognitive extension, as well as the differences between external and internal cognitive processes, Sutton (2010) draws on Merlin Donald's (1991) explanation of "exograms"—external symbols—and how these are comparable (but different) to engrams—the traces and changes in the brain that occur during memory formation. The process of using external structures in, say, the process of remembering, allows us to "create and support cognitive profiles quite unlike those of creatures restricted to the brain's biological memories and engrams alone" (Sutton, 2010, p. 189).[20] But this does not assume that the internal processes and external tools are functionally the same. To recap, internalists argue that: (1) internal and external processes should be treated as two different kinds because they are realized in different ways; and (2) external vehicles do not process information in 'genuinely' cognitive ways because there is too much variety in how they process information (Adams & Aizawa, 2001, 2006; Rupert, 2004). However, for Sutton (2010) and Menary (2006), as well as Clark (2008), these criticisms are largely based on a misreading of the parity principle as an argument in and of itself. Such a reading is mistaken because, as Clark (2008) notes, the parity principle should only be viewed as a probe to "free ourselves from bio-chauvinistic influences" (p. 114). When speaking of the parity principle as a 'rule of thumb' Clark (2008) states:

> In other words, for the purposes of identifying the material vehicles of cognitive states and processes, we should (normatively speaking) ignore the old metabolic boundaries of skin and skull and attend to the computational and functional organization of the problem-solving whole. (p. 77)

Menary (2006) echoes these sentiments further: instead of acting as a fully formed argument for extended cognition, the parity principle is meant as an "intuition pump" designed to steer us away from "Cartesian prejudices" (p. 333). That aside, Sutton (2010) admits that "exclusive

[20] An example of how actors remember plays can be found in §1.6 below.

focus" on cases of parity "can cause trouble" (p. 196). This is, in part, because the parity principle "does not encourage attention to the distinct features of the components in particular extended cognitive systems" (Sutton, 2010, p. 198). As noted earlier, it is this downplaying of differences, due to the parity principle's functionalist commitments, that is most problematic. We therefore require additional clarification while also addressing the issues raised by internalists.

The complementarity principle can offer us solutions. Returning to engrams and exograms, Sutton (2010) highlights some differences, as pointed out by Donald (1991). Exograms are more enduring and have greater informational capacity than engrams, and they can be better manipulated and are more transmissible across "media and context" (Sutton, 2010, p189). But, despite these differences, it would be a mistake to consider cognitive artifacts, like exograms, as "simply commodities" used by the brain to profit—i.e., only used to achieve cognitive success in memory tasks, for example (Sutton, 2010, p190). Rather, they actually can become a part of cognitive processes.[21] Sutton (2010), drawing on Clark (2003), elaborates:

> The human mind is "leaky" both because it thus extends beyond the skin to co-opt external devices, technologies, and other people, and because our plastic brains naturally soak up labels, inner objects, and representational schemes, internalizing and incorporating such resources and often redeploying them in novel ways. (p. 190).

Along these lines, the complementarity principle may be compatible with the parity principle, but we should focus on complementarity since it offers "more naturalistic answers to objections" (Sutton, 2010, p. 194). We are thus moving away from functionalism and closer to connectionism. As Sutton (2010) explains:

> With this complementarity principle, as we might call it, we return connectionism to the heart of the case for EM. It's just because isolated items aren't stored atomically in the brain that our relatively vulnerable biological memories are supplemented by more stable external scaffolding. Brains like ours need media, objects, and other people to function fully as minds. (p. 205).

It is this 'connectionism' that must be highlighted, and this is the core of explaining our coupling with deep scaffolding.

Traditional connectionism posits that neural networks best describe how certain cognitive processes occur (Buckner & Garson, 2019). A connectionist system involves

[21] Sutton (2010) says that, in certain circumstances, these cognitive artifacts are "(a part) of the mind" (p. 190). I have changed this to "a part of cognitive processes" to avoid inflationary extended mind claims.

"networks of nodes that excite or inhibit each other's activity according to weighted connections between them" (Shapiro & Spaulding, 2021). Embodied cognition—the idea that the body and its interactions with the environment significantly impact cognition—has incorporated many connectionist principles (Clark, 1997). For example, some hexapod robots (robot insects, as it were) are controlled by a neural net and walk by responding to direct feedback from features of their environment, such as the floor (Beer & Chiel, 1993, as cited in Clark, 1997). When walking across uneven terrain, if a lowered leg's sensors detect no purchase, the leg is drawn back and placed in a different location until suitable ground is found (Clark, 1997, p. 17). Other robots achieve walking by "detecting the forces exerted by the terrain so as to compensate for slopes" (Clark, 1997, p. 15). In these instances, walking is achieved through sensory feedback with motor processes, allowing the robots to adjust their walking appropriately. When we walk, the feedback loops involved are similar: processes start in brain and link to the feet (which touch the floor). The sensory feedback then goes back into the body, to the brain, and allows us to compensate for certain surfaces, such as slippery ones.[22] These systems allow new situations (and feedback) to cause different movement adaptations (Clark, 1997, p. 92). In the complementarity principle, cognitive processes also link with external structures and illustrate a looping and coupling—which is a vital component in the completion of certain cognitive tasks. The walking embodied robot and human cases are not 'deep', but they parallel how deep scaffolding establishes complementarity with internal cognitive systems. The interaction and feedback between internal and external systems creates a new cognitive process—even though the external structures do not perform the same functions as the internal systems. Returning to our manipulation of a slide rule, the scaffolding allows us to perform mathematical calculations that would otherwise not be possible with internal resources alone. Under Sutton's (2010) view, external media, in some form or another, is therefore required in order for our brains to function fully. This follows Clark's (2003) core argument in *Natural-Born Cyborgs*: it is in our nature as human beings to be "human-technology symbionts" who spread thinking, reasoning and even ourselves across the "biological brain and nonbiological circuitry" (p. 3). The complementarity principle can account for our coupling with external structures in ways that the parity principle (alone) cannot. When we use a slide rule, it is not that the artefact is performing an internally equivalent process, but rather the coupling allows a new cognitive process to be performed. In this way, the external component deserves cognitive status due to its significant involvement. By using the complementarity principle, we can now bypass the shortcomings of the parity principle. By accepting that internal

[22] Clark (1997) describes how these adaptations are not the result of learning but are "inherent in the original dynamics of the system" (p. 92).

and external vehicles are different, even though we are giving up on some functionalist commitments, we can answer the internalist's first objection. The complementarity principle treats internal and external vehicles as functionally and constitutionally different things, which is vital in arguing for integration (as we saw with engrams and exograms).

The connectionism point can be elaborated further. Menary (2006) gives his own explanation for how internal and external vehicles perform joint cognitive processes. He lays out four theses that the complementarity picture is committed to: the manipulation thesis, the hybrid thesis, the transformation thesis and the cognitive norms thesis (Menary, 2006). Let us begin with the manipulation thesis. This considers the ways that agents perform cognitive tasks by manipulating vehicles in their environment. This can be an individual process, involving one agent, or a co-operative one, involving multiple agents. Menary (2006) identifies three types of manipulation. These include:

> 1. Biological cases of coupling such as extended phenotypes and animate vision (biological coupling).
>
> 2. Using the environment as its own representation, obviating the need for internal representations—as in Tetris (i.e., epistemic actions).
>
> 3. And most importantly, the manipulation of external representational and notational systems according to certain normative practices—as in mathematics (i.e., cognitive practices). (Menary, 2006, p. 331)

Type 1 manipulations can be seen in how biological couplings take place between organisms. For example, a male cricket's song is a biological coupling when it directs female crickets to the song's origin (Webb, 1994, as cited in Menary, 2006). These are typically shallow but could also be deep, as discussed with an ant colony's pheromones. We saw many instances of type 2 manipulations in §1.3's discussion of epistemic actions and these are ubiquitous in the extended cognition literature—overlapping with shallow cases (such as object reminders) and deep cases (such as slide rules). And lastly, type 3 manipulations, involve the use of cognitive norms when manipulating external systems—as also demonstrated with slide rules, written mathematical equations and calendars, to name a few—and are particularly characteristic of deep scaffolding[23]. Next, Menary's (2006) hybrid thesis again emphasizes connectionist principles. It specifies that cognition integrates non-classical internal vehicles and processes (such as neural networks) with classical external vehicles and processes. Menary (2006) also points to how certain instances of

[23] For example, when an agent uses pen and paper to solve long division problems.

remembering can be considered products of the internal and external vehicles which "complement each other in the completion of a cognitive task" (p. 330). The next thesis, the transformation thesis, states that our cognitive effectiveness is transformed by "learning the *practice* of manipulating external vehicles" (Menary, 2006, p. 330). The successful transformation of cognitive processes thereby requires the appropriate use of scaffolding and practices. A navigational map, for example, needs to be rigid with regards to how it displays certain locations, and we have to apply the appropriate cognitive practices to successfully navigate—such as using the legend or scale. The transformation thesis is another way of describing deep scaffolding, since the appropriate manipulation of external vehicles to complete cognitive tasks is considered a cognitive process. Lastly, closely linked with the transformation thesis, the cognitive norms thesis makes explicit the point that cognitive integration not only depends on internal and external vehicles and processes, but also the correct application of cognitive norms and practices (Menary, 2006).[24] We are able to perform long division using a notepad and pen when we correctly learn and apply "the norms by which we manipulate mathematical notations" (Menary, 2006, p. 330). These cognitive norms and practices "are both cognitive and distinguishable from other norms and practices because they are aimed at the completion of cognitive tasks" (Menary, 2006, p. 330).

Now that these theses have been laid out, let us apply them to a deep scaffolding example: our trusty calendar. Consider a busy hospital that needs to solve the problem of determining which doctors can and cannot be on leave. In this instance, no more than two doctors can be on leave at the same time. Using a calendar and coloured stickers, each doctor places stickers on their leave days. This then transforms a hard computational problem into an easier one by allowing each doctor to determine when they can take time off. So, how are Menary's theses being applied. According to the manipulation thesis, we are using type 3 manipulations to manipulate an external representation of the month (i.e., the calendar). The doctors then use the stickers to represent their leave days and the calendar is used communally. The hybrid thesis demonstrates the cognitive integration of internal vehicles with the calendar system. Doctor A needs to determine whether he can go on leave from the 18th to 28th of October. He then sees that Doctor B has placed stickers from the 15th to the 29th meaning that his proposed days are not viable. The transformation and cognitive norms thesis then show us that the doctors' cognitive effectiveness for determining their leave days is transformed by them learning to manipulate the calendar and applying the appropriate cognitive norms and practices

[24] We saw this with the map example.

when placing stickers. We can thus view the connectionism involved in cases of complementarity as follows: X (agent) + Y (external structure) allow X to perform Z (a cognitive process, such as remembering) (Menary, 2006, pp. 333-334).[25] X coupled with Y (deep scaffolding, such as a calendar) allows X to perform the cognitive processes of Z (determining which days X can go on leave). This coupling allows deep scaffolding to be given cognitive status when used in the completion of cognitive tasks (by virtue of being an important part of the cognitive system). We can therefore call cases of deep scaffolding cognitive extension. Additionally, we can also give a more detailed explanation for bypassing the internalist's second objection—that external vehicles do not process information in genuinely cognitive ways—because complementarity argues that internal and external vehicles are parts of a whole cognitive system, which allows certain cognitive processes to be completed. As previously noted, the parity principle acts only as a way to start qualifying certain external processes as cognitive (Clark, 2008). The complementarity principle then gives us the explanation for how cognitive extension of this kind is possible.

There remains, however, one more significant objection to be addressed: Adams and Aizawa's (2006) coupling constitution fallacy. Adams and Aizawa (2006) (henceforth referred to as A&A) argue that only traditional internal cognition can be constituted as intrinsically cognitive. Otto's notebook, for example, is an artifact that is coupled with a cognitive system, but the notebook itself is not intrinsically cognitive in the same way that brains are (Adams & Aizawa, 2006). A&A (2006) argue that "the causal coupling of X with Y does not make X a part of Y" (Menary, 2006, p. 333). However, as we discussed above, Menary (2006) posits that this view is predicated on A&A (2006) using the parity principle as a fully formed argument in and of itself. To counter this, we should reject the Otto example on the basis of the parity principle alone, since it implies "a discrete, already, formed cognitive agent" before even coming into contact with an external structure (Menary, 2006, p. 333). The complementarity principle then buttresses the parity principle, primarily by appealing to the manipulation and hybrid theses. By using the manipulation thesis—which notes that agents manipulating vehicles in their environment to perform cognitive processes—and the hybrid thesis—which notes that cognition integrates both internal and external vehicles and processes—we should instead view certain cognitive processes (such as remembering in the Otto case) as only being possible when a notebook is involved (Menary, 2006). As Menary (2006) lays out:

> X is the manipulation of the notebook reciprocally coupled to Y—the brain processes—which together constitute Z, the process of remembering. Once we have this picture, it is

[25] A similar coupling is found in all of our previously discussed deep scaffolding cases.

> easy to see that A&A have distorted the aim of cognitive integration. The aim is not to show that artifacts get to be part of cognition just because they are causally coupled to a pre-existing cognitive agent, but to explain why X and Y are so coordinated that they together function as Z, which causes further behavior. (p. 334)

Manipulating external vehicles thereby plays an inherent role in the completion of certain cognitive tasks and our couplings deep scaffolding work in this way. Under the complementarity principle, when we use calendars and slide rules, the external components are so integral to the completion of cognitive tasks that the internal and external processes should be considered parts of a unified system. This forms new cognitive processes that either make existing processes easier—such as the calendar that determines when doctors can go on leave—or allows for the completion of processes that are (likely) not possible with internal resources alone—such as performing complex mathematical problems with pen and paper. The external vehicles involved then deserve a share of the epistemic credit and deep scaffolding is therefore genuinely cognitive.[26]

## 1.6. Sterelny's Dimensions of Scaffolding

In this section, I consider Sterelny's (2010) three scaffolding dimensions (trust, entrenchment and sharing). These dimensions further our understanding of how agents develop complementary relationships with scaffolding—most notably in deep cases. I have already used the complementarity principle to show how cognitive extension is possible (i.e., through the coupling of internal and external vehicles). However, there are other important factors regarding an agent's coupling with scaffolding that warrant attention. Drawing from Peter Godfrey-Smith's (2009, as cited in Sterelny, 2010) model for developing a multi-dimensional characterisation of evolutionary regimes, Sterelny (2010) lays out three dimensions for scaffolding. I summarize these, following Piredda (2019), as trust, entrenchment and sharing[27]. These dimensions are not always strict, much like our 'deep' and 'shallow' distinction, and some overlap can occur. Sterelny's (2010) original purpose for describing these dimensions was to show that extended

[26] Shallow scaffolding is not cognitive in the same way. While certainly a part of the explanatory story for how an agent's cognitive processes operate, it has far fewer explanatory demands than deep cases. With shallow scaffolding, complementarity between the agent and scaffolding is minimal. This is because, when certain organisms react in cue-bound ways, the cues simply result in agents responding in particular ways.

[27] Sterelny (2010) also refers to this third dimension as 'The individual and the collective'. He uses it as a broader point to show how we should not consider communal resources through the extended mind model (Sterelny, 2010). I will not enter that debate here, so I am instead drawing more from Piredda's (2019) summary where she refers to this dimension as "sharing" (p. 14).

mind cases are "limiting special cases of scaffolded minds" (p. 473). Although acknowledging complementarity and cognitive integration, this seems to adhere more to the first wave view of extended mind.[28] But rather than enter that discussion here, I use Sterelny's (2010) dimensions to demonstrate how we develop strong couplings with scaffolding—which has implications for the following chapters.[29]

Let us begin with the dimension of trust. Sterelny (2010) states that "trust involves the agent's assessment of the reliability of their access to a resource and the reliability of the resource itself" (p. 473). He also adds that shared resources are sometimes trustworthy because they are shared (Sterelny, 2010, p. 474). A subway map, for example, is seen as trustworthy because many different agents use it reliably. This is important for explaining an agent's coupling with scaffolding because they first establish trust with an external resource before optimal complementarity is formed. If we believed that a map was doctored in some way, trust would be broken and complementarity lost. Deep scaffolds, especially, require this dimension of trust before they can be used successfully in the completion of cognitive tasks. Trust is also important in shallow scaffolding; but in cases where agents are less reflexive, it is easier to bypass. Ants 'trust' their pheromone signals even when they are being deceived— as demonstrated by certain beetle species who mimic ant chemical signatures to infiltrate colonies (Hölldobler & Wilson, 1990, pp. 498-505). If we, however, imagine that a stop sign is misplaced to deceive drivers, even though some may stop, others would be able to use additional cues or prior knowledge to realize the deception.

The next dimension is entrenchment. This includes what Sterelny (2010) refers to as interchangeability and individualisation. Let us first consider supports that are interchangeable. Sterelny (2010) imagines a generic set of cooking knifes. These can be used interchangeably because their use is standardised. Given correct usage, most will be able to succeed in carrying out a specific cutting task. When we apply this idea of interchangeability to scaffolding, we can point to road lane markings, slide rules and emergency signs. These, too, are standardised for general use. Once the appropriate norms are learnt, they assist most agents in the same way. That said, let us now consider, as Sterelny (2010) does, a professional cook's knives. If we were to examine them, we would see that they are individualised to suit the preferences of the agent (Sterelny, 2010, p. 475). The same may be said of a professional batsman's equipment or a blind person's walking stick (Sterelny, 2010). In these cases, complementarity is improved and the

[28] §1.5. argued that deep scaffolding should be viewed under the second wave of version of extended mind.
[29] These implications include how hostile scaffolding might bypass or take advantage of trust and entrenchment.

objects may even begin to feel like extensions of the agents' hands when used (Sterelny, 2010, p. 477). However, once these supports have been modified, they also become more difficult for other agents to use—reducing complementarity. The objects cease to be interchangeable and are now individualised. By individualising a tool or scaffold, we entrench it deeper into to an agent's cognitive functions. In this way, "the relationship between agent and environmental support develops over time" (Sterelny, 2010, p. 476). A customised notebook or calendar system could fit this build as well, as they "are used persistently because they have been individualised" (Sterelny, 2010, p. 476). Entrenched resources can also include other agents, such as a mother reducing her child's degrees of freedom so they can perform certain tasks.[30] Furthermore, we also saw entrenchment in affective artefacts. Piredda (2019) refers to these as personal affective artefacts and indicates that they are also more reliable and regularly used.[31] These can influence an agent's affective states with a certain degree of regularity (and the relationship develops over a time), as seen with a wedding ring or Linus' blanket. These "interact with the affective condition of the agent with a certain regularity over a certain period of time, and are thus the most interesting and typical ones" (Piredda, 2019, p. 6). A secondary feature of these artefacts is that they have the potential to extended or enhance our sense of self. These objects may therefore include an aspect of "self-extension or self-resonance – the feeling that our self is somehow extended or enhanced through these objects" (Piredda, 2019, p. 7). So, overall, the more entrenched scaffolding becomes, the more reliable it is for completing certain cognitive tasks. Scaffolding, be it deep or shallow, seems to adhere to this dimension of entrenchment. And, in the case of deep scaffolding, the more we individualise an environmental support, the more we increase its complementarity with our own cognitive systems.

The final dimension, sharing, overlaps with trust. This examines how scaffolds are either used exclusively by an individual, or by a community. Sterelny (2010) looks at Evelyn Tribble's (2005) analysis of environmental cognitive supports in the early modern Globe Theatre. In this case:

> The scripts and the physical layout of the theatre itself were organised to constrain the range of possibilities and to cue action at the right moments. Plays themselves had a somewhat stereotyped overall organization, and the stage had a stable organization through which action was funnelled in highly constrained ways. Likewise, scripts

[30] Sterelny (2010) argues that we would find it difficult to say that the mother is an extension of the child's mind (as in extended mind cases), because she cannot be individualised in the same way as artefacts. But this is, again, seems to follow on from a first wave extended mind understanding. The mother being a part of the child's cognitive process is much less problematic under the complementary principle.

[31] A random blanket may not offer the same sense of security as his blanket.

provided actors with skeletal information on how a given play varied from the predictable flow of entry and exit. (Sterelny, 2010, p. 477)

What we see here are shared communal resources which cue and reduce the agents' degrees of freedom when performing tasks (in this case, performing a play).[32] As Sterelny (2010) previously noted under the trust dimension, sometimes resources are trusted specifically because they are shared (and inherited). In the Globe Theatre example, agents are "acting jointly and collaboratively" as well as taking "advantage of a space they have in part inherited from others" (Sterelny, 2010, p. 477). On the surface, this may indicate that inherited communal (shared) resources are more trustworthy to agents, since they appear to work reliably and many agents use them successfully.

Sterelny's (2010) dimensions are useful in understanding how scaffolding (and, most notably, deep scaffolding) can establish complementarity with our cognitive systems. Trust must first be established with a scaffold if it is to be used effectively, and entrenchment deepens our complementarity with a resource. Additionally, the dimension of sharing seems to emphasise that communally resources are often more trustworthy to agents. If scaffolding were to, in principle, be hostile, then it would have to bypass the dimension of trust. If that then became entrenched in an agent's cognitive system, its effectiveness in undermining an agent's interests would be increased. Furthermore, just because a resource is shared, that does not necessarily limit its potential to have negative effects—as will be discussed in Chapter 3.

## 1.7. Conclusion: Turning Hostile

We can summarize the important moves of Chapter 1 as: (1) laying out key scaffolding thinkers, (2) expanding on our characterisation of scaffolding by drawing a distinction between shallow and deep scaffolding, (3) appealing to the complementarity principle to explain deep scaffolding, and (4) further explaining how we establish trustworthy and entrenched couplings with scaffolding. Throughout these sections, a common theme can be observed: there is a focus on scaffolding in a 'benign' sense (Clark 1997, 2003; Hutchins, 1995a; Piredda, 2019; Sterelny, 2010). That is to say that the scaffolds usually benefit the agents using them. Benign scaffolding can

[32] Sterelny (2010) notes that collectively achieved knowledge, norms and practices do not fit directly under an extended mind model (such as Otto's notebook) because of how shared, multi-layered and complex they are (p. 478). I am, instead, focusing on how this dimension divides scaffolding into individual and collective resources.

therefore be characterised as: external structures that change the cognitive demands of some tasks in ways that serve the interests of the agent.

Clark (1997, 2003), for example, tends to highlight the great benefits of scaffolding in his work. He often echoes how "it is the human brain plus these chunks of external scaffolding that finally constitutes the smart, rational inference engine we call mind" (Clark, 1997, p. 180). But the less positive effects of scaffolding are often not considered. Clark (1997) states that some structures "allow us to impute 'preferences' on the basis of the constraints of success in… a larger machine" (p. 183). Depending on the degree of density and control, a highly scaffolded environment can directly impact how freely individual agents can act within it. In the case of a heavily scaffolded firm or organization, individual rationality may become "somewhat marginal" (Clark, 1997, p. 186). An organisation's goals are often determined through a "quasi-evolutionary selection of constraining policies and institutional practices" (Clark, 1997, p. 182). If the primary goal is to maximize efficiency and profit, for example, then tasks may be geared towards that end and individual agents could become more akin to cogs in a machine. In such cases, it could be possible for an organisation's goals to conflict with the welfare of agents (Slaby, 2016). Here is an instance where scaffolding may not always be beneficial and it could undermine some agents' interests, while serving those of others. But these particular implications are not discussed by Clark (1997).

Sterelny (2010), on the other hand, acknowledges some of the dangers of scaffolding and notes that it may be possible for someone to benefit in certain "one-on-one high-stakes negotiations" (p. 474). For example, an agent may be able to tamper with the information inside Otto's notebook to serve their own ends (Sterelny, 2010). But while the "hostile manipulation of [our] informational environment is a serious danger", Sterelny (2010) restricts this to a limited set of "single-sender, single-receiver systems" (p. 474). For Sterelny (2010), while it may be possible in the Otto case, it is unlikely that an agent could get away with manipulating shared resources. Suppose someone changed all the maps in a subway station for deceptive purposes (Sterelny, 2010). In this case, Sterelny (2010) argues that it would likely be very difficult to successfully manipulate agents in specific ways, the returns would be little when compared to the costs and other agents could collectively verify that something was wrong. Sterelny (2010) states that:

> Informational resources in a shared space are sometimes reliable because they are shared. Such resources are the joint product of many agents and are typically used at unpredictable times and places. In many circumstances, public domain resources cannot safely be used to manipulate a specific target for a specific purpose (p. 474).

But such a view may de-emphasise the extent to which shared scaffolds are manipulable and vulnerable to hostility, such as scaffolding in casino environments (which will be discussed in §3.3).

Framing the discussion around scaffolding in a (mostly) benign way is understandable. If we look back at the conventional definition of scaffolding—a structure that facilitates other tasks, such as construction—the connotations are positive. Likewise, in our historical overview (§1.2), we saw that 'scaffolding' was widely used in the context of learning and skill development (Shvarts & Bakker, 2019). Jan Slaby (2016) notes these positive approaches are based on 'user/resource' accounts: where agents are only seen to use external structures as beneficial tools. This approach may have limited our considerations of scaffolding that is hostile towards agents. In the next chapter, I will argue this possibility is far greater than what is portrayed. If we are engaged in complex cognitive feedback loops with external structures, and we know that the environment is often hostile (i.e., serving the interests of other agents) then scaffolding must also be vulnerable to hostility, irrespective of its shared or private nature. 'Hostile scaffolding' therefore refers to external structures that change the cognitive demands of a task in ways that undermine the interests of one agent, while serving the interests of another agent. Furthermore, if scaffolding can be deep, then 'deep hostile scaffolding' could also exist—which would involve the manipulation of agents' cognitive systems.

# Chapter 2: Making the Case for Hostile Scaffolding

## 2.1. Introduction

At the end of Chapter 1, I gestured towards the neglected possibility of 'hostile scaffolding': external structures that change the cognitive demands of tasks in ways that undermine the interests of one agent, while serving those of another agent. If we are engaged in complementary cognitive feedback loops with external structures, and we know that the environment is often hostile (serving the interests of competitors/ predators), then hostility could also be expressed through these structures. However, rather than giving this type of scaffolding significant attention, a great deal of the literature views scaffolding as benign (or neutral) to agents' interests (Clark, 1997; Hollan et al., 2000; Sterelny, 2010). This chapter aims to develop the view that scaffolding can be hostile while acknowledging other thinkers who discuss similar environmental harms (Aagaard, 2020; Liao & Huebner, 2020; Slaby, 2016).

My use of the word 'hostile' draws from Sterelny's (2003) notion of hostility in *Thought in a Hostile World*. Sterelny (2003) suggests that environments are not always informationally transparent; they can also be informationally opaque and hostile. I begin by discussing these views (§2.2) before clarifying how I am handling 'interests' (§2.3). This involves examining what is meant by an agent's 'local' interests as well as their 'actual' (or 'overall') interests. I then argue that we can differentiate cases of hostile scaffolding from cases of harmful scaffolding by identifying the agent who suffers and the agent who benefits. The beneficiary is required to be causally relevant— 'in the loop', as it were—for the scaffolding to be hostile. Next, I re-emphasise that, because humans are distinguished from other animals by how much of our activity is scaffolded (Clark, 2003; Hutchins, 1995b; Kirsh, 1996; Sterelny, 2010), our heavy use of scaffolding can also create unusual vulnerabilities (§2.4). The possibility of hostile scaffolding in our cognitive lives is therefore a genuine concern. The final aim of this chapter is to examine other relevant works (§2.5). I discuss: Jesper Aagaard's (2020) *4E cognition and the dogma of harmony*, Jan Slaby's (2016) paper *Mind Invasion*, as well as Shen-yi Liao and Bryce Huebner's (2020) *Oppressive Things* (2020). While these do touch on two or three of the Es of 4E cognition, they do not quite get to the idea of hostile (or deep hostile) scaffolding. I then conclude Chapter 2 by setting up Chapter 3—which will provide real world cases of hostile scaffolding (§2.6).

## 2.2. Characterising 'Hostility'

Although not explicitly stated, Sterelny's (2003) book, *Thought in a Hostile World*, can be read as an argument against the claim that "the world is its own best model" (Brooks, 1990, Section 3, para. 1; see also Brooks, 1991). This is because environments are not always informationally transparent—they can also be informationally opaque and hostile (Sterelny, 2003). I begin by using Sterelny's (2003) work to specify what is meant by my use of the term 'hostility'.

Just as agents can compete over resources, they can compete over information as well. An informationally transparent environment can be characterised by "simple and reliable correspondences between sensory cues and functional properties" (Sterelny, 2003, p. 17). Cue-driven agents will often succeed when the cues generated by their environment (and other agents) are reliable and honest. Many species of migrant birds, for example, use celestial objects to successfully guide navigation, and ants of the same colony may signal (through pheromones) that they share common nest membership (Sterelny, 2003, p. 21). However, transparency is not always the case and environments can also be informationally opaque and hostile. While detection systems—which Sterelny refers to as "single-cued discriminatory mechanisms"—are often useful in allowing organisms to respond appropriately to their environment, they can also be exploited by others (Sterelny, 2003, p. 24). For example, camouflage and mimicry are some of the ways in which animals make environments more informationally opaque. Recall, in §1.3, we discussed ant scaffolding in the form of pheromone trails. These reduced the cognitive demands of tasks—such as locating food resources and identifying group members. However, despite its useful applications, Sterelny (2003) notes how an ant colonies' pheromone system is also open to exploitation:

> Ants recognize and react to one another by specific chemical and mechanical cues, so parasites bearing no physical or other resemblance to the ants can invade and exploit their nests by mimicking the right specific signals. Thus a number of beetle species live in ant nests, mimicking the ants' chemical signature and the foodbegging gestures. These beetles persuade their hosts to feed them and even to tolerate their feeding upon the ants' larvae. (p. 15)

Here, we have a hostile case: scaffolding that usually benefits some agents (the ants) but allows agents of another type (the beetle species) to manipulate specific cues in order to profit. A similar example can be found in female *Photuris* fireflies—aggressive mimics that feed on another species of male fireflies by using light signals to attract them (Lloyd, 1997). Because most fireflies use (helpful) cues for finding suitable mates, *Photuris* fireflies can mimic those cues and benefit. In the above cases, "camouflage and mimicry make signals less reliable, and they make signals that are still reliable harder to discriminate" (Sterelny, 2003, p. 25). By producing and exploiting certain cues, the beetles and fireflies can undermine the interests of agents that are (usually) served by scaffolding. Such cases are examples of shallow hostile scaffolding due to the agents' reliance on cues. The world is therefore not (always reliably) its own best model, since informational hostility and deception can be core features of an agent's environment.

The above hostile cases mainly focus on the exploitation of detection systems—which are (single cue) discriminatory mechanisms that mediate an organism's adaptive responses to particular environmental signals (Sterelny, 2003, p. 14). Single cue-driven agents can face difficulties when cues map onto physical signals in complex ways (Sterelny, 2003, p. 21). This may make certain tasks, such as distinguishing mate from predator, more difficult. The greater the quantity of signals (both honest and hostile), the more an agent's epistemic environment is polluted (Sterelny, 2003, p. 25). From an evolutionarily perspective, increasingly complex and hostile environments may have led to the development of more robust tracking systems: those which involve multiple cues (Sterelny, 2003, pp. 27-29). Female reed warblers, for example, have to navigate the problem of their nests being invaded by cuckoo eggs. An invariant strategy would be detrimental, since a reed warbler that treated all of her eggs equally would always fall victim to brood parasites (Sterelny, 2003, p. 11). Reed warblers therefore have to use more than single cues to detect cuckoo eggs in their nests. These include recognizing egg size and colour, the timing of eggs being placed into nests, and the sighting of cuckoos in the area (Sterelny, 2003, p. 28). These additional cues mean that reed warblers are more likely to defend their own reproductive interests. When it comes to humans, Sterelny (2003) discusses how we may have developed more advanced cognitive systems in the form of decoupled representations. According to this view, "we have internal cognitive states which (a) function to track features of the environment, and (b) are not tightly coupled functionally to specific types of response" (Sterelny, 2003, pp. 30-31). This allows us a greater degree of reflexiveness. That said, while robust tracking and decoupled representations generally better serve agents when responding to informationally hostile and complex environments, at least when compared to single-cue detections systems, they do not fully bypass the issue. Robust tracking does not result in all cuckoo eggs being rejected and errors can always occur. And, with regard to decoupled representations, "no mechanism is perfectly reliable, and hence not all representations will be accurate" (Sterelny, 2003, p. 31). Hostility remains a large concern that all types of cognitive systems are vulnerable to.

Possible errors and inaccuracies aside, the cuckoo egg example may indicate how hostile scaffolding has also had to develop alongside other cognitive systems. As reed-warblers have evolved more ways to detect cuckoo eggs, so too have cuckoos evolved more ways to bypass those systems. Cuckoos lay eggs fast, when reed warblers are away foraging, and their eggs have become more mimetic of other species' eggs (Sterelny, 2003, p. 28). This could suggest that, the more robust a cognitive system is, the harder hostile scaffolding has to work to reliably bypass it—especially in shallow cases. But, if the complementarity approach in §1.5 is correct, then deep hostile scaffolding is a far more subtle way of bypassing cognitive systems: the scaffolding actually becomes a part of an agent's cognitive processing (Menary, 2006; Sutton, 2010). So, if we could identify genuine examples of deep hostile scaffolding, they would be more interesting than shallow cases due to how they are integrated with cognitive processes. Cuckoo eggs that succeed in deceiving other bird species, despite needing to bypass multiple-cues, are still more shallow than they are deep. It would be difficult to say that the cuckoo egg that is more mimetic of a reed warbler egg is actually facilitating significant cognitive processing. The shape, size and colour of the egg is just one of many cues that the reed warbler uses to identify potential imposters. The same can be said of other single cue scaffolds, such as the ant pheromone and firefly examples. By labelling these cases 'shallow hostile scaffolding', we are just using new terminology to describe deception, mimicry and camouflage. This new vocabulary is still useful to develop, as we will see in §3.2 and §3.3.1, but identifying (and expanding on) deep hostile scaffolding cases highlights a currently underdeveloped phenomenon.[33]

So, we can review the main takeaways from Sterelny (2003) as follows: environments are not just informationally transparent but also epistemically and informationally hostile. We can characterise 'hostility' as a property of an environment, relative to an agent, that undermines the interests of certain agents, while serving those of others. Hostile scaffolding then applies this idea to external structures that change the cognitive demands of a task by undermining the interests of the agent attempting the task, while serving the interests of another agent. Shallow hostile scaffolding can already be observed in the ant, firefly and cuckoo examples, since they rely on the exploitation of cue-based systems. But this simply re-describes deception, mimicry and camouflage. It is therefore more interesting to attempt to identify and explain cases of deep hostile scaffolding.

[33] Identifying deep cases would also allow us to view hostile scaffolding under a 4E cognition lens.

## 2.3. Interests: "Who Suffers?" & "Who Benefits?"

Because my characterisation of hostility involves the undermining and serving of interests, this interests criterion calls for some additional clarification. Most thinkers who assume that scaffolding is (mostly) benign are not required to address interests this way since, in those instances, the agent is usually assisted by scaffolding (Clark, 1997; Hutchins, 1995a; Sterelny, 2010). For this project, however, my characterisations of benign and hostile scaffolding refer to interests, which need to be spelled out more carefully. So, what do we mean by an agent's interests being served or undermined? Can scaffolding be harmful without being hostile? Is scaffolding that undermines an agent's interests but benefits no one hostile? What if an agent's interests are undermined but the other agent benefits randomly or by accident? This section aims to provide answers. First, in order to talk about an agent's interests being subverted or benefitted, we need some kind of theory of interests. Scaffolding serves an agent's interests when it allows them to successfully perform a task and is in line with their 'actual' (or 'overall') interests—i.e., what is good for an agent's flourishing. Second, we need a way to separate hostile scaffolding from harmful scaffolding—scaffolding which harms an agent simply because it is ineffective at achieving certain goals. And third, in the case of hostile scaffolding, we need to specify that the agent who benefits does not do so by accident; rather, it is through their actions that they benefit. The beneficiary therefore needs to be causally relevant and 'in the loop'. Before continuing, it should be noted that the theory of interests I develop here will not be comprehensive and some will disagree. I, however, aim to do enough, briefly and clearly, to answer the above questions while developing criteria for identifying hostile scaffolding cases. A full, comprehensive theory of interests is not needed to accomplish this task.

Let us start by addressing the first question: What do we mean by an agent's interests being served or undermined? Throughout this section, I will be using terms like 'local', 'actual' and 'overall' interests. Local interests are relative to the task being attempted and are linked to success conditions. For example, you may desire a calm state of mind and decide to listen to classical music to achieve that goal. The success of the scaffolding (i.e., whether the music serves your local interests) can be determined relative to the success conditions of achieving a calm state. Matters become more complicated, however, when considering an agent's actual (or overall) interests. By actual interests, I mean interests that are actually 'good' for the agent. These are linked to the agent's welfare, and can conflict with local interests and tasks—such as an addict seeking a fix. Imagine a gambling addict who has the local interest of gambling. By being inside a casino environment and using gambling machines, they may be able to successfully achieve their local interest but, in doing so, they may be unable to pay for food or rent. We could say that, in this instance, being able to afford food and shelter is more beneficial to the agent's health than gambling. Alternatively, the agent may have the local interest of gambling but the overall interest of no longer being addicted. The pulls of the casino environment could then override this overall interest (Schüll, 2012). In both gambling cases, the agent's local interests are served while their actual interests are subverted. In some situations, it may be difficult to determine whether an agent's actual interests are being subverted by their local interests. For example, what if the previously discussed calming music distracts you from other more important work? I will not address this in detail here. But it needs to be acknowledged that the actual interests of an agent are sometimes unclear. However, in cases like the gambling addict, it is safe to assume that the agent's interest to gamble is often in conflict with their actual interests.[34]

The above discussion also raises another important question: Can scaffolding be harmful without being hostile? Recall the object reminders discussed in §1.4. These scaffolds, such as a strategically placed bottle of olive oil, can cue additional cognitive functions, like remembering. Now consider a similar example, but this time substitute the olive oil bottle with a wine bottle. If Joseph wants to remember to buy more wine (his local interest), we can ask if the strategically placed bottle cued his memory appropriately. If we answer yes, then it would seem that the scaffolding is successful and the object reminder is benign relative to Joseph's local interest. But, much like the gambler, Joseph's interests can be in conflict with one another. He may want to remember to buy a new bottle of wine but, if he is an alcoholic, this could negatively affect his health, work and family life. In this case, the object reminder is harmful, but it does not directly serve another agent's interests either. Or, alternatively, we can imagine a scenario where Joseph is not an alcoholic but the object causes him mental distress by serving as a reminder of his alcoholic father. The scaffolding is, again, potentially doing more harmful than good, but no one seems to directly benefit. I will refer to such cases as 'harmful scaffolding': external structures that change the cognitive demands of tasks but fail to (or ineffectively) serve the interests of the agent performing the task—while also not benefitting another (causally relevant) agent. This is demonstrated in the wine bottle cases but could take to form of badly designed scaffolding—such as a convoluted slide-rule that increased the cognitive demands of a task. The benign and hostile distinction may, therefore, not exhaust the options for the forms that scaffolding can

[34] This is made clear by numerous interviews conducted by Natasha Schüll (2012) in her book, *Addiction by Design*. Many gambling addicts describe the "ever-present awareness of being in a destructive process", yet they are unable to extricate themselves from their situation (Katrina, as cited in Schüll, 2003, p. 24).

take.[35] That said, I will not discuss these possibilities in greater detail since hostile scaffolding remains my primary focus.

But the issue of harmful scaffolding does highlight our third and fourth questions: Is scaffolding that undermines an agent's interests but benefits no one still hostile, and what if the other agent benefits randomly or by accident? As laid out in §2.2 above, a core feature of hostility is that an agent's interests are undermined while another agent benefits (Sterelny, 2003). Harmful and hostile scaffolding both subvert an agent's interests but we can distinguish between them by asking "Who suffers?" (*Cui malo?*) and "Who benefits?" (*Cui bono?*)[36]. In both Joseph cases, we can identify Joseph as the agent who suffers, since his overall interests are harmed by using the scaffolding. But it would be problematic to say that no one benefits, since there could always be unwitting beneficiaries. The question then becomes: What if the other agent benefits randomly or by accident? Imagine that, because of the bottle reminders, Joseph's alcoholism worsens, resulting in him being fired from his job. Jess, an employee working directly under Joseph, is then promoted to Joseph's former position. Here, Joseph's competitor, seemingly benefits by accident. It is clear that such a case should not qualify as hostile scaffolding (since the other agent is not 'directly' involved) and we therefore need to stipulate that the beneficiary is causally 'in the loop'—i.e., they have causal relevance by contributing to the presence of hostile features. The beneficiary then benefits directly from the ways in which the victim is undermined.[37] Being causally 'in the loop' also fits with the other hostile cases, such as the *Photuris* fireflies—since deceptive light signals causally influence other species' behaviour (Lloyd, 1997). Because Jess is not causally relevant (she benefits accidentally), this is a case of harmful scaffolding and not hostile scaffolding. But we could picture a similar scenario that is hostile. Imagine that Jess knows of Joseph's alcoholism and breaks into his apartment every night to set empty bottles on the kitchen counter, knowing that they will cue Joseph to buy more wine. If these object reminders worsened Joseph's alcoholism, resulting in his firing, then Jess is now both the beneficiary and causally 'in the loop'. Although contrived and unlikely, this example now gives us a (shallow) hostile scaffolding case. Let us ask "Who suffers?" and "Who benefits?" in another example: the gambling addict. If we ask "C*ui malo?*", we can identify our gambler as the agent who suffers. If we ask "C*ui bono?*", we can identify the casino as financially benefiting.

[35] Cases like the convoluted slide-rule could be referred to as 'functionally incompetent scaffolding'—scaffolding which is badly designed—while the Joseph cases could be called 'agent-relative harmful scaffolding'—scaffolding which usually serves most agents' interests but does not serve a particular agent's interests.
[36] My use of "*Cui bono?"* is inspired by Daniel Dennett (2006), who noted that the *cui bono?* question should be asked when discussing reasons for any agent's behaviour.
[37] Without this qualifier, we may identify a disproportionate amount of hostile scaffolding cases by confusing them with harmful cases.

Furthermore, the casino is causally 'in the loop' because they work with designers to create scaffolding with the stated purpose of promoting gambling activity (Friedman, 2000; Schüll, 2012). This provides us a real-world case of hostile scaffolding (and will be explored further in Chapter 3).

In summary, we now have a clearer understanding about what is meant by 'interests' and I have demonstrated that local and actual (overall) interests can be in conflict with each other. Most importantly, when analysing scaffolding cases, we now have criteria for identifying hostile scaffolding. By asking "Who suffers?" and "Who benefits?", as well as determining whether or not the beneficiary is causally 'in the loop', we can successfully separate hostile scaffolding from harmful scaffolding. We can then determine if the hostile scaffolding is shallow or deep by highlighting if the scaffolding employs cues or facilitates significant cognitive work.

## 2.4. Humans & Scaffolding: Reliance & Vulnerability

Now that we can better identify hostile scaffolding cases, it is important reiterate the degree to which we are potentially made vulnerable to them. As discussed in Chapter 1, various scholars have examined how external structures play significant roles in our cognitive lives. This list includes Clark (1997, 2003), Sterelny (2010), Hutchins (1995a, 1995b) and Kirsh (1996). By re-emphasising their views, I argue that hostile scaffolding could reveal the existence of an unusual vulnerability: because we use cognitive scaffolding so extensively, we could also be exploited by hostile external structures. I begin by briefly reviewing Kirsh (1996) and Hutchins (1995a), before discussing Sterelny (2010) and Clark (2003). As we review each thinker, I also explain how hostility (as laid out in §2.2 above) is not addressed.

Kirsh (1996) discusses how agents, particularly humans, can adapt environments to themselves by making them more 'cognitively hospitable'. He calls the measurement of how 'cognitively hospitable' an environment is its "*cognitive congeniality*" (Kirsh, 1996, p. 440). Epistemic actions[38] and complementary actions[39] all allow us to improve *cognitive congeniality* (Kirsh, 1996). Kirsh (1996), too, stresses that we find these types of actions everywhere in human life (p. 449). Similarly, Hutchins' (1995a, 1995b) views are also important to re-emphasise. In Chapter 1, I discussed how ships' navigational systems could be viewed as their own cognitive systems that

[38] As noted in Chapter 1, when we perform epistemic actions, such as placing a key in a shoe or organising workspaces, we scaffold our environments to more easily perform cognitive tasks (Kirsh & Maglio ,1994).
[39] These are "an interleaved sequence of mental and physical actions that results in a problem being solved" (Kirsh, 1996, p. 442).

could be 'stepped inside' (Hutchins, 1995a). Hutchins (1995b) takes a comparable approach when analysing airline cockpits. Speed bugs, for example, display air speed information and are just one of several instruments that aid in memory tasks. Speed card booklets then act as the cockpit's long-term system and "[store] a set of correspondences between weights and speeds that are functionally durable"—which allow appropriate speeds to be chosen (Hutchins, 1995b, p. 276). Weights do not have to be remembered, since they are stored on cards, and the organisation of the crew further distributes the cognitive burden of tasks. The successful completion of piloting duties should therefore be attributed to these socio-technical systems (rather than to individual minds alone) (Hutchins, 1995b). But despite analysing the degree to which these structures and environments make us cognitively successful, Hutchins (1995a, 1995b) and Kirsh (1996) do not address how they could make us vulnerable. While agents may be able to make some tasks more viable by improving an environment's *cognitive congeniality*, others could use external structures to muddy *cognitive congeniality* and profit (as shown in §2.2). Furthermore, creating cognitive systems that we can 'step inside' may create additional opportunities for cognitive manipulation (which could occur in deep hostile scaffolding cases).

Shifting now to Sterelny (2010), as discussed in §1.3, he argues that we can best explain our cognitive capacities by the ways our cognition is supported by the environment—noting that our "inner mechanisms have coevolved with and adapted to this rich environment" (p. 471). The scaffolded mind hypothesis then "proposes that human cognitive capacities both depend on and have been transformed by environmental resources" (Sterelny, 2010, p. 472). Such a view can be contrasted with Sterelny's (2003) earlier work, discussed in §2.2 above, where he characterises environments as also having the capacity to be informationally hostile. It is, therefore, surprising that Sterelny (2010) does not focus more on scaffolding's capacity to be hostile—outside of a set of limited interactions. As noted in §1.7, these included the hostile manipulation of informational resources in 'one-on-one high-stakes negotiations'—such as an agent tampering with Otto's notebook—and the manipulation of shared resources—such as an agent changing all the subway maps in an area (Sterelny, 2010). This latter case was seen as particularly unlikely. Sterelny also discusses the potential real-world implementation of Monty Python's Hungarian-English phrasebook (Sterelny, 2010, p. 474). In the well-known sketch, Hungarian users unwittingly make absurd and sometimes offensive utterances—such as "my hovercraft is full of eels"—when attempting to communicate in English. But Sterelny states that "it is difficult to envisage circumstances in which an author would gain from producing a maliciously misleading phrasebook, for an author cannot know when, where, by whom or with what effect such a book will be read" (Sterelny, 2010, p. 474). Even though Sterelny (2010) is correct when considering these examples, restricting hostile scaffolding to these limited interactions may undersell the extent to which hostile scaffolding is present in our everyday lives.[40] Furthermore, some shared resources may still be vulnerable to hostility despite their shared nature.[41]

Finally, let us return to Clark. §1.3 discussed Clark's (1997) *Being There* and drew on a number of examples.[42] Clark (2003) would later expand on these views in *Natural-Born Cyborgs* by emphasising how our cognitive lives have always been defined by the use of external tools. Clarks (2003) states:

> It is our special character, as human beings, to be forever driven to create, co-opt, annex, and exploit nonbiological props and scaffoldings. We have been designed, by Mother Nature, to exploit deep neural plasticity in order to become one with our best and most reliable tools. (pp. 6-7).

Although we may one day become the cyborgs of science fiction literature (i.e., physically modifying our bodies and cognitive processes with technology), Clark (2003) argues that we have always been 'cyborgs' (in a sense), due to the ways we have co-opted external tools to complete cognitive tasks. To demonstrate the degree to which we are aided by scaffolding, Clark (2003) asks us to imagine a workday routine where we have to deliver a meeting (pp. 25-28). In summary: First, you are woken by your alarm clock (rather than your circadian rhythm) at 07:30 am. On your way to the office, despite driving over a section of icy road, your car's traction control and ABS kick in to prevent skidding. When you arrive, you begin working by making use of a large file, which includes your previous drafts as well as other colleagues' work. While consulting the file, your internal cognitive systems are able to add additional ideas and notes. You then use your laptop to view more saved information, coaxing your biological brain to, again, "respond with a few fragmentary hints and suggestions" (Clark, 2003, p. 26). Lastly, you bring a summarized version of key points, collected from your files and notes, into the meeting and successfully address your audience. For Clark (2003), assuming that your biological brain (alone) is responsible for the successful completion of these tasks is misleading. The story of how we accomplish certain goals requires scaffolding to be a part of the process. As Clark (2003) states:

> What the human brain is best at is learning to be a team player in a problem-solving field populated by an incredible variety of nonbiological props, scaffoldings, instruments, and

[40] A point which Chapter 3 addresses in more detail.
[41] Chapter 3 also discusses how casino environments can bypass multiple reliable agents using shared resources.
[42] These included the use of pen and paper for long division as well as instances of socio-institutional scaffolding (Clark, 1997).

resources. In this way ours are essentially the brains of natural-born cyborgs, ever-eager to dovetail their activity to the increasingly complex technological envelopes in which they develop, mature, and operate. (p. 26)

As Clark (2003) (and others[43]) indicate, surrounding ourselves with scaffolding is often crucial to how we problem-solve and think. It can be argued, along the lines of Sutton (2010) and Menary (2006), that this scaffolding is an integral part of our cognitive systems.[44] Our reliance on scaffolding is therefore characteristic of our cognitive lives. But, because our environment can also be hostile (serving the interests of competitors/ predators), hostility could 'piggyback' on top of external structures and into our cognitive lives.

Clark (2003) does acknowledge that our use of scaffolding can create vulnerabilities, but his emphasis is mostly optimistic. He considers the following areas: inequality, intrusion, uncontrollability, overload, alienation, narrowing, disembodiment, deceit and degradation (Clark, 2003, pp. 167-195). Let us briefly summarize and touch on the unrelated concerns first. 'Inequality' addresses how wealth disparity and education could divide our ability to use technology. 'Intrusion' considers how technology (such as Internet cookies) can unwittingly intrude into our private lives and compromise our data. 'Uncontrollability' addresses the worry that the more we rely on human-machine symbiosis, the more our freedom and 'humanity' will be lost. 'Overload' refers to the sheer quantity of stimuli (such as messages and e-mails) we may be exposed to. 'Alienation' refers to how agent technologies (like smart chat bots or semi-intelligent interfaces) might obviate the need for genuine human connection by mimicking it. 'Narrowing' considers how software can limit what is recommended to agents, thus narrowing the media they are exposed to. And, 'disembodiment' refers to: (1) how an agent who overuses technology may feel a greater connection with their online presence and avatars than their own body; or (2) how, with an ever-increasing focus on technological data transmission, the human body may be seen as less necessary (Clark, 2003). These concerns may qualify as harmful scaffolding in some cases and, while very real and relevant, do not consider the degree to which scaffolding itself could be used by other agents to profit.[45]

Clark's (2003) concerns of 'deceit' and 'degradation' are the closest he gets to addressing informational hostility (and opacity). 'Deceit' refers to how the Internet allows devious

[43] This includes this section's previously discussed authors: Hutchins (1995a, 1995b), Kirsh (1996) and Sterelny (2010).
[44] Clark (2003) also specifically refers to a "*complementarity*" process (p. 75).
[45] Clark's (2003) considers how 'intrusion' (as seen in Internet cookies) could benefit other agents (such as advertisers) but his focus is not on how those external structures change cognitive demands of tasks.

individuals to more easily advance their own interests—by lying about their identities, for example. Clark (2003) discusses a case where white supremacists posed as African-Americans (advocating outrageous claims, such as the legalization of paedophilia) in an attempt to smear other racial groups (p. 183). Similarly, issues such as 'catfishing'—where a fictional online persona lures victims into relationships (often sexual)—and cyber-bots—bots that pose as actual people[46]—are also discussed (Clark, 2003, pp. 184-185). 'Degradation', on the other hand, is a related concern regarding a lack of quality control (Clark, 2003). With the Internet making it easier for agents to send and publish information, a different type of 'overload' occurs, as it becomes increasingly difficult to "separate the wheat from the chaff" (Clark, 2003, p. 187). In these cases, we can make comparisons with Sterelny's (2003) informational opaqueness and hostility. The Internet has indeed created an environment where reliable signals are more difficult to detect and where deceitful agents can more easily exploit others. But informational opaqueness is an issue all agents have to solve; the Internet is just a different environment where these challenges have to be met.

Clark's (2003) concerns regarding 'deceit' do present us with cases of hostile scaffolding—external structures that change the cognitive demands of tasks in ways that undermine the interests of an agent, while benefitting those of another. But these cases remain in the domain of shallow hostile scaffolding. Just as *Photuris* fireflies mimics light signals, so too do human agents mimic the online personas of different individuals. Furthermore, these concerns do not take a 4E cognition approach—such as the consideration of deep hostile scaffolding. In this way, Clark (2003) does not consider how our extended cognitive loops might be subverted. Furthermore, his responses remain optimistic. These include the development and refinement of technologies like CAPTCHA to combat cyber-bots and the balancing of misinformation with legitimate 'truths' (Clark, 2003, pp. 185-186). It is also suggested that online services take more responsibility for the information they publish—punishing those who post false information and rewarding those who are historically shown to be trustworthy (Clark, 2003, p. 189). These responses, while helpful, do not offer solutions that acknowledge the degree to which our nature[47] (a point which Clark (1997, 2003) emphasises) is partly what could make us vulnerable to cognitive exploitation.

In this section, I re-emphasised the degree to which we use external structures in our cognitive lives. Despite various scholars (Clark, 2003; Hutchins, 1995a, 1995b; Kirsh, 1996;

[46] These bots may, for example, pose as voters or participants in surveys (Clark, 2003, p. 184).
[47] By this, I mean our nature to co-opt external structures to complete cognitive tasks (Clark 1997, 2003; Sterelny, 2010).

Sterelny, 2010) discussing our ubiquitous use of scaffolding, the ways in which this could make us vulnerable to hostility have not been properly considered. This can partly be account for by how these authors have differing goals than those of mine. Hutchins (1995a, 1995b) developed a theoretical framework to discuss distributed cognition. Kirsh (1996) discussed how *cognitive congeniality* can be improved through epistemic and complementary actions. Sterelny's (2010) main goal in *Minds: extended or scaffolded?* was to demonstrate that extended mind cases, like Otto, are limiting cases of environmental scaffolding. And Clark (2003) tried to sell us on the idea that external structures are an inseparable part of our mental lives—i.e., that we are, in some sense, already 'cyborgs'. The optimism regarding our co-opting of external resources stems from these foci. But it remains the case that negative instances of scaffolding are not fully considered and that the potential harms (resulting from our cognitive reliance on scaffolds) are underemphasised. The next section will discuss thinkers who take a different approach—demonstrating that scaffolding can sometimes harm agents.

## 2.5. Other Related Work: 4E Cognition & Environmental Harms

Some thinkers acknowledge that scaffolding can be helpful in some ways but negative in others. In this section, I discuss Aagaard's (2020) paper *4E cognition and the dogma of harmony*, Slaby's (2016) *Mind Invasion*, and Liao and Huebner's (2020) *Oppressive Things*. These works all examine environmental harms through a 4E cognition lens, where at least two or three of the 4Es are discussed. But, despite parallels with my project, the idea of hostile scaffolding is not quite reached. As we discuss each paper, these parallels will be shown. I argue that my research's focus on hostility may offer new ways to discuss certain types of situated environmental harms.

### *The Dogma of Harmony*

Some scholars acknowledge the benefits of a 4E approach to cognition while also being critical of it—in a similar manner to my points raised in §1.7. Aagaard (2020) has also identified the tendency for the discourse to focus on external structures and their positive cognitive effects. He dubs this rhetoric the 'dogma of harmony' and describes it as "an overly idealized picture of human–technology relations in which all entities are presumed to cooperate and collaborate" (Aagaard, 2020, p. 2). More specifically, Aagaard (2020) believes that the 'dogma of harmony' neglects the "problems associated with technology use" (p. 8). The two main phenomena highlighted are 'deskilling' and 'bad habits'.

Let us begin with 'bad habits'. Aagaard (2020) notes that Internet browsing technologies have created many bad habits by reducing our self-control and attention spans. Rather than focus on the negative aspects of Internet browsers, some authors have chosen to highlight more beneficial programs, such as 'StayFocused' (Clowes, 2019, as cited in Aagaard, 2020). These programs proport to improve concentration by closing browser windows after specific periods of time—encouraging the formation of good habits over bad ones. Similarly, the program 'Freedom' limits a user's access to Wi-Fi until their computer is rebooted (Dotson, 2012, as cited in Aagaard, 2020). However, as Aagaard (2020) points out, these approaches neglect how browser technologies have exacerbated issues of self-control in the first place. The need for these beneficial technologies challenges "the idea that Internet technologies enhance human agency in any straightforward way" (Aagaard, 2020, p. 12). Aagaard (2020) argues that Internet browsers, and smartphones, have the potential to induce 'digital akrasia'[48] when they "make us act in ways that go against personal values such as being attentive and present during everyday conversations" (p. 12). When looking at our hostile scaffolding criteria, we see that the agent who suffers has already been identified by Aagaard (2020) (the person using a device with browser software). However, Aagaard (2020) does not address the agent who benefits. We could ask: Can 'digital akrasia' benefit platforms like Google? If the answer is "yes", and we can place Google causally 'in the loop', we may have a hostile scaffolding case. While discussing 'digital akrasia' in isolation is a worthwhile project, hostile scaffolding could add additional considerations by addressing how browser technology can impede on an agent's self-control while benefitting the interests of other agents (like platforms and advertisers).

Aagaard (2020) also discusses 'de-skilling' and argues that 'outsourcing' our cognitive resources is not always beneficial—it can sometimes diminish our own cognitive capacities or make us passive (p. 13). Examples include how smart cooking appliances (such as a smart toaster) may hinder the development of good cooking skills; or how Global Positioning Systems (GPS) may automatise navigation in ways that erode our own navigation skills—leading to diminished agency (Aagaard, 2020, p. 13). Discussing this 'deskilling' as it relates to hostile scaffolding may be difficult, since it is not clear that 'deskilling' is simply an unavoidable consequence of certain 'on-board' cognitive processes being made unnecessary. At worst, this may be harmful scaffolding. However, Aagaard's main point still stands: Despite the benefits of cognitive supports, we could also be weakened by being supported. Even though this view does

[48] Aagaard (2020) defines 'digital akrasia' as "the tendency to become swept up by one's digital devices in spite of better intentions" (p. 12).

not (explicitly) state that we are then exploited after being cognitively weakened (which could benefit other agents), the possibility of hostility is, at least, compatible with this project.

Despite these potential compatibilities, Aagaard (2020) does not quite get to the idea of hostile scaffolding. Among his suggestions for avoiding the 'dogma of harmony', he proposes that we should broaden the scope of 4E cognition to include both positive and negative cases, and that we should not presume that all entities involved in cognitive extension are working towards a shared goal (Aagaard, 2020, p. 14). Developing criteria for hostile scaffolding and listing examples may allow us to broaden the discussion in precisely those ways.

### *Mind Invasion*

Other attempts to argue that scaffolding is not always benign include Slaby's (2016) paper, *Mind Invasion*. Slaby (2016) delves into how a social organization's norms and goals can, rather than aid cognition, hijack certain affective states. Also drawing on Clark (1997), Slaby (2016) describes an external scaffold as "any item or structure in the environment that provides reliable support for cognitive processes" (p. 4). He primarily focuses on situated affectivity and argues that certain environments can 'invade' the minds of their employees by altering their values and emotional/motivational states (Slaby, 2016, p. 11). While mainly considering corporate workspaces, Slaby (2016) indicates that 'mind invasion' is also possible in higher education, social-web-based subcultures, sports, as well as the security, military and police sectors. Much like Clark's (1997, 2003) views, Slaby (2016) argues that even though we can shape these scaffolded environments, they are also important in shaping us (p. 2). They are, in this way, often "prior to and formative of individual emotion repertoires and affective-bodily styles" (Slaby, 2016, p.2).

But Slaby (2016) differs from Clark (1997) (and others) by disagreeing with their optimistic focus. One potential explanation for the prevalence of benign scaffolding is a 'problematic' which Slaby (2016) calls "the predominance of the 'user/resource model'" (p. 7). Slaby (2016) summarizes this as follows:

> Baseline mentality in many of the example cases under discussion is that of a fully conscious individual cognizer ('user') who sets about pursuing a well-defined task through intentional employment of a piece of equipment or exploitation of an environmental structure ('resource'). (p. 7)

Such an account sees a (non-conflicted) agent who is fully aware of the details of a task, as well as scaffolding that is benign (or neutral) to the welfare of the agent. Slaby (2016) posits that, because proponents of situated cognition tend to favour this approach, they have "failed to acknowledge the potentially troublesome political issues that the situatedness perspective might make visible" (p. 7). In this way, Slaby's (2016) critique is partly political, since he discusses how certain structuring effects can exploit workers. He also highlights Clark's (1997, 2003, 2008) approach as being emblematic of an overemphasis on the 'user/resource model' (a view which I share). Similarly, it is argued that Colombetti and Krueger (2015), as well as Griffiths and Scarantino (2009), also gravitate (or at least hint) towards scaffolding as useful tools that achieve desired affective states. I made similar connections with the thinkers in §2.4 above and we can also include the positive rhetoric used in Menary (2006) and Sutton's (2010) work, since they described external and internal vehicles that jointly perform helpful cognitive processes. Slaby (2016) emphasises that, in most of these cases:

> the individual with his or her interests, inclinations, intentions and strategies is taken for granted as a starting point that is then placed in purposeful conjunction with a technical device or an environmental structure so that an effective coupled system of 'user-plus-tool' results (p. 8).

It could be that the optimists Slaby (2016) describes are guilty of assuming that local interests are mostly compatible with overall interests. The term 'mind invasion' attempts to express the ways in which some external structures can be affectively pervading when they influence users in ways that are not beneficial to their well-being—i.e., in ways that undermine their overall interests. Similarly, Slaby (2016) highlights a second issue: the tendency for some thinkers to be "unwilling to sufficiently distinguish between a process-oriented and a normative understanding of its subject matter" (p. 9). The exclusion of a normative dimension[49], has resulted in some neglecting how socio-normative dimensions can "enable and constrain individual mental states" (Slaby, 2016, p. 9). Mind extension approaches thereby often exclude social frameworks when considering the mind and cognition.

Let us now consider how Slaby's (2016) 'mind invasion' works. Work domains with particularly strong 'cultures' have the tendency to create coherence and conformity with those cultures. In a corporate setting, a newly hired recruit will be subjected to a great deal of affective influences. The ways in which other employees work, interact, conduct themselves, address superiors and even use humour can create influential affective niches (Slaby, 2016). Despite initial feelings of discomfort or confusion, the new employee may (over time) find themselves

[49] By normative dimension, Slaby (2016) is referring to individual mental states being "more like public moves in a rule-governed game – or like the commitments and entitlements accrued to the games' players in virtue of their moves" (p. 9). These then constrain and enable certain mental states.

habituated to that environment. New recruits will then develop new working styles, goals and ways of approaching social and work situations based on the affective niches they were exposed to. Throughout this process of affective shaping, they may be socially rewarded or punished when they conform, or fail to conform, with certain comportments. This becomes problematic when these new styles go "discernibly against these individuals' prior orientations, and if it is in the long run detrimental to their personal flourishing" (Slaby, 2016, p. 2). As Slaby (2016) notes, a simple example can be found in how an unanswered email can "weigh upon one, exerting a subtle affective pressure until one finally goes about answering" (p. 15). Similarly, having an active email account may result in us constantly refreshing our inbox for fear of "losing track or being left out of relevant procedures at work" (Slaby, 2016, p. 15). Shutting down your computer or smartphone may lead to similar feelings of worry and distress. This issue is compounded by the phenomenon of 'presence bleed' (Gregg, 2011, as cited in Slaby, 2016). As the means to take work out of the office become more advanced and pervasive, with smartphones and constant email accesses, working hours can also increase. In such cases, the boundaries of working hours and off-hours may begin to fade. Slaby (2016) argues that this can result in the exploitation of employees, whereby a company can influence an agent's affective states to extract the most value (p. 14). These practices can then undermine an employee's actual (overall) interests.

Slaby's (2016) work is useful for explaining why the 'user/resource model' has been (mostly) taken for granted. Optimistic approaches assume that scaffolding is benign (or neutral) to an agents' welfare and that users are a fully conscious cognizers with well-defined tasks. Slaby (2016) shows that, instead of taking this at face value, we should consider how external structures can negatively influence an agent by shifting their values in ways that are detrimental to their well-being (i.e., through 'mind invasion'). I argue that Slaby is correct here. That said, the notion of 'hostility' (as sketched in §2.2 above) is missing from Slaby's picture. This is understandable since we saw how Slaby's (2016) critique is partly political—focusing on the "subjectification effects of social domains" and the exploitation of workers (p. 2). Additionally, Slaby (2016) focuses on situated affectivity, whereas my project includes affectivity as well as traditional cognitive processes. And lastly, he does not discuss whether this scaffolding could be deep.[50] But despite these differences, the idea of 'mind invasion' is compatible with hostile scaffolding and my project can expand on Slaby's (2016) examples. Consider a laptop that is always logged on to a company's email server. The company also promotes a culture of long workdays as well as weekend work—which results in employees constantly checking their email,

[50] This last point maybe be contentious for affective scaffolds.

even during off-hours. Here, we can identify the agent who suffers as an employee (when the resulting affective states harm their well-being); and we can identify the beneficiary as the employer (or company) (who benefits from increased employee productivity). Finally, the employer could be causally 'in the loop' if they actively promote unhealthy, work centric values (or the use of technology that leads to 'presence bleed'). This would be an instance of hostile scaffolding (or, at least, harmful scaffolding).

### *Oppressive Things*

The last related work I discuss is Liao and Huebner's (2020) *Oppressive Things*. Here, Liao and Huebner (2020) discuss artefacts and environments that can be hostile towards a group of agents. They argue that both spatial environments and material objects can be racist when congruent with oppressive systems (Liao & Huebner, 2020).[51] By congruent, they mean: objects and environments which are "*biased in the same direction* as other manifestations of an oppressive system"; are "*causally embedded* in the respective oppressive system"; and have bi-directional causal connections due to being products of these systems while also "guid[ing] and constrain[ing] racist psychological processes and racist social structures" (Liao & Huebner, 2020, p. 9). More broadly, Liao and Huebner (2020) draw on an externalist view of racism when discussing 'racist things'. Under this view, "racism is not to be found just in individuals' minds (or their bones), but in the ways that individuals' interact with other individuals and social institutions" (Liao & Huebner, 2020, pp. 6-7). Because 'racism' involves extrinsic relationships, the term can be applied to objects and environments as well.

Liao and Huebner (2020) discuss the example of Kodak's Shirley Cards, originally implemented in the 1950s (Del Barco, 2014). During the process of photo printing, a Shirley Card was used to calibrate skin-color balance. As Liao & Huebner (2020) explain:

> A photo of a "Shirley"—so named after the model in the original incarnation—was typically printed first; and so long as Shirley's ivory skin, brown hair, and red lipstick looked "right", photographers could expect subsequent printing to look "right" as well. (p. 3)

Liao and Huebner (2020) note that, even though it may have been necessary for Shirley Cards to be used, given certain technological limitations, they are still racist objects when taken in the context of an oppressive system. The cards only displayed one type of natural skin colour to be

[51] Liao and Huebner (2020) characterise oppression as unjust relations where people are compelled to adjust their cognition to a world that "rarely centers their interests" (p. 6).

printed correctly and this also prescribed normality—making aesthetic and normative judgements on how 'natural skin colour' should appear. In contrast, dark skin tones would come out as "over saturated or underlit" (Liao & Huebner, 2020, p. 4). Such entrenched racism privileges "the descriptive norms encoded by the objects" (Liao & Huebner, 2020, p. 5). Furthermore, the decisions taken by engineers to use one type of technology over another, even if that technology presented legitimate technical limitations, could be the result of additional light-skin biases (Liao & Huebner, 2020). Arguments against such stances usually ignore the influence of the prescriptive norms from which various techniques are chosen (Liao & Huebner, 2020, p. 5). Objects like Shirley Cards can serve as "anchors for social practices" and changing them requires "transforming the broader cultural background" (Fields & Fields, 2014, as cited in Liao & Huebner, 2020, p. 5).

Shirley Cards fit our scaffolding discussion, since they changed the cognitive demands of tasks—allowing Kodak employees to adjust colour calibration by comparing an initial print with the card. But these scaffolds also harm agents by promoting racist beliefs and norms. If we focus on the bi-directional nature of Shirley Cards, they are external structures that changed the cognitive demands of colour calibration while also acting as material anchors[52] for racism (based on the normative value they promoted) (Liao & Huebner, 2020). Racist things and environments thereby "partially constitute the stability and structure of this racial frame" while shaping "habit of attention and categorization and the attitudes" of agents within an oppressive system (Liao & Huebner, 2020, p. 7). When viewed under a broader network of racist objects and environments, we see the development of oppressive cognitive niches—which can also be applied to ableist and sexist things and environments as well (Liao & Huebner, 2020, p. 9).

I argue that Liao and Huebner's (2020) are correct about racist objects and environments. These subvert the interests of oppressed agents (by perpetuating racist normative values and frameworks) and may be congruent with hostile scaffolding cases. We could identify the agents who suffer as members of the oppressed group; and we could identify the beneficiaries as members of the oppressor group. A Kodak employee using a Shirley Card is using cues to verify colour calibration, but the scaffolding is also presenting a normative assessment of how skin colour should look.[53] Much like Slaby (2016), this is a broader focus than that of my project, with

[52] Liao and Huebner (2020) state that material anchors "are aspects of the physical world, which generate intrapersonal and interpersonal forms of stability in social spaces, by shaping patterns of association, behavior, and imagining" (Hutchins, 2005, as cited in Liao & Huebner, 2020, pp. 10-11).
[53] It is also possible that the Shirley Card is doing more cognitive processing than just cuing the agent. The 'deep' structure could be the larger racist culture in which the card and agents are embedded. But such considerations are presently outside the scope of this dissertation.

Liao and Huebner (2020) focusing on 'hostility' as it relates to a group of people (in an oppressive system). This dissertation localises the discussion to how an individual's interests are undermined while attempting certain cognitive tasks. That said, the ways in which this project is compatible with Liao and Huebner's (2020) work requires more expansion. This could include addressing, more specifically, how interests are served and undermined; how oppressor agents are causally 'in the loop'; as well as deep and shallow considerations. (A related future project is briefly discussed §4.1.) But for now, given the scope of this dissertation, localising my focus to individuals allows the parameters of the interests being served/ undermined to be more easily defined. Additionally, Liao and Huebner (2020) accept two of the 4Es (embedded and embodied cognition) and state that they are 'sympathetic' to enacted cognition (p. 3). What they do not accept is the first-wave conception of extended cognition, although they do not make a judgement on the second-wave (Liao & Huebner, 2020). My project is centred around the second-wave of extended cognition which, I argue, is compatible with Liao and Huebner's (2020) views.

## 2.6. Conclusion

This chapter made some important moves in setting up Chapter 3's hostile scaffolding cases. I began by clarifying what is meant by the term 'hostility'—a property of an environment, relative to an agent, that undermines the interests of one agent, while serving those of another (Sterelny, 2003). I then elaborated on 'interests' and developed criteria for identifying hostile scaffolding cases. We can distinguish hostile scaffolding from harmful scaffolding by asking "Who suffers?", "Who benefits?" and whether or not the beneficiary is casually 'in the loop'. I then discussed Hutchins (1995b), Kirsh (1996), Sterelny (2010) and Clark (2003) to re-emphasised how humans make heavy use of scaffolding when completing cognitive tasks—which can open us up to negative effects as well. Despite that possibility, I demonstrated how this is not widely discussed. I then closed by reviewing thinkers who acknowledged that external structure can sometimes be harmful (Aagaard, 2020; Liao & Huebner, 2020; Slaby, 2016). However, hostility, as characterised in §2.2, was not described—as well as the possibility of deep hostile scaffolding. In Chapter 3, I apply the criteria developed in this chapter to show that hostile scaffolding is a genuine concern and, most importantly, that some cases are deep.

# Chapter 3: Hostile Scaffolding in the Wild

## 3.1. Introduction

In Chapter 1, I expanded on scaffolding by drawing a distinction between deep and shallow cases. Recall that shallow scaffolding involves agents using cues to reduce the cognitive demands of tasks, whereas deep scaffolding involves external structures offloading significant portions of cognitive processing. Next, in Chapter 2, I characterised 'hostile scaffolding' as external structures that change the cognitive demands of tasks by undermining the interests of one agent while benefitting those of another. Now, in Chapter 3, I will provide hostile scaffolding examples of both types.

While some shallow hostile cases have already been discussed—such as certain beetle species that mimic ant pheromones (§2.2)—I begin this chapter by providing additional examples. These include sunglasses (Viola, 2022) (§3.2) and casino interior design and ambience (Friedman, 2000; Schüll, 2012) (§3.3.1). Such cases highlight the presence of hostile scaffolding in everyday life while also serving as warmup exercises[54]. Real examples of deep hostile scaffolding, on the other hand, are theoretically possible but have so far remained elusive. I present three deep hostile examples: (1) gambling machines that utilise player tracking (Schüll, 2012) (§3.3.2), (2) slot-machines that use 'virtual reel mapping' (Schüll, 2012) (§3.3.2), and (3) Twitter's gamification of communication (Nguyen, 2021) (§3.4). As we discuss these examples, I use the hostile scaffolding criteria (developed in §2.3) to properly classify each case. I identify: the agent who suffers, the agent who benefits, the task, the external structure(s) (ES), how ES transforms the cognitive demands of a task, the harms for agent 1, the benefits for agent 2, whether or not agent 2 is causally 'in the loop', and whether the scaffolding is shallow or deep.

## 3.2. Sunglasses as Hostile Scaffolding

Let us begin with our first warmup case: sunglasses as hostile scaffolding. Setting aside the practical use of protecting our eyes from glaring sunlight, Marco Viola (2022) argues that sunglasses are a type of affective scaffolding. They can act as social shields and display emotional self-control to other agents by allowing wearers to hide negative emotions (Viola, 2022, p. 3). I

[54] These will show how we can begin using the hostile scaffolding criteria laid out in Chapter 2.

argue that sunglasses can become shallow hostile scaffolding when agents gain by obfuscating certain emotional tells.

Some facial expressions can be: (a) "spontaneous expressions of emotion" while others can be (b) "aimed at producing some effect within a social interaction" (Viola, 2022, pp. 5-6). While some agents may use (b) type expressions to produce desired social effects, (a) type expressions can 'leak' emotional information that one may wish to keep private (Viola, 2022, p. 6). In these later cases, sunglasses act as scaffolding when they transform how easily other agents can identifying facial patterns—thus preventing 'leaks'. Poker players, for example, sometimes wear sunglasses to disguise involuntary 'tells' after their hand is dealt. And Viola (2022) cites studies (Kim et al., 2022; Noyes et al., 2021, as cited in Viola, 2022) that confirm the common-sense notion that "sunglasses confuse the perception of some emotions, especially sadness and fear" (p. 8).

Let us imagine a potentially hostile case. Suppose you are purchasing a used car from and a salesperson wearing sunglasses. As you ask questions regarding previous owners and the vehicle's condition, you find it more difficult to read her facial expressions—at least when compared to the non-sunglasses wearing salespeople. Her answers all stretch the truth but you fail to detect any facial expressions that indicate otherwise. You eventually decide to purchase the car believing it to be a good faith transaction. Now imagine that the same interaction takes place but this time without the sunglasses. As your questions are answered, you detect a worried glance here or a furrowed brow there. You get an uneasy feeling and, this time, decide not to purchase the vehicle. The previous interaction (where you made the purchase) is an instance of shallow hostile scaffolding. Of course, there are other factors that reveal deception and the mere presence of the sunglasses may be reason enough to doubt the salesperson's authenticity. But this example shows that, when sunglasses transform the cognitive demands of interpersonal interactions, they could serve the interests of the wearer while undermining another agent's interests. Here, the task is to determine the salesperson's trustworthiness by reading their facial expressions. The sunglasses change the cognitive demand of this task—making them scaffolding. The scaffolding changes the cognitive demands by making it difficult for agent 1 (you) to determine agent 2's (the salesperson's) facial expressions. We can identify the agent that suffers as agent 1—if you are successfully deceived—while agent 2 is the beneficiary—i.e., the salesperson financially benefits if they succeed in deceiving you. This is also hostile rather than harmful scaffolding because agent 2 is causally 'in the loop'— the salesperson intentionally wears the sunglasses to conceal emotional 'leaks'. Finally, this is a shallow case because the sunglasses obfuscate facial cues rather than offloading significant processing work.

We can conclude our sunglasses case by summarizing the key points in a table:

| | |
|---|---|
| Agent 1 (Who suffers?): | The agent interacting with the agent wearing sunglasses. |
| Agent 2 (Who benefits?): | The agent wearing sunglasses. |
| The task: | To make sense of interpersonal interactions. |
| The external structure(s) (ES): | The sunglasses. |
| Does ES transform the cognitive demands of a task? | Yes. The sunglasses make it difficult for agent 1 to determine agent 2's facial expressions. |
| Does ES harm agent 1? | Yes. Agent 1 is unable to accurately read agent 2's facial expressions and is deceived. |
| Does ES benefit agent 2? | Yes. Agent 2 financially benefits by deceiving agent 1. |
| Is Agent 2 causally 'in the loop'? | Yes. Agent 2 intentionally wears the sunglasses to conceal emotional 'leaks'. |
| Shallow or deep? | Shallow. The sunglasses obfuscate facial cues. |

In the above table, we can observe that the criteria for shallow hostile scaffolding have been met. Sunglasses can therefore be hostile scaffolding in situations where obfuscating facial expressions benefits the wearer and deceives other agents.

## 3.3. Hostility & Scaffolding in Las Vegas

Some casinos use both shallow and deep hostile scaffolding to undermine player interests.[55] In this section, I primarily draw form Natasha Schüll's (2012) book, *Addiction by Design*, which offers a superb overview of the gambling industry in Las Vegas. Among other things, Schüll (2012) draws attention to the victims of machine gambling and the techniques used by casinos and game designers to keep players spending. This presents a deep well of hostile scaffolding cases to draw from.

Some players intentionally use gambling machines as powerful methods of "affective self-management" (Schüll, 2012, p. 19). Throughout this section, I will sometimes refer to a player flow state called 'the zone'. Mollie, a gambling addict, describes 'the zone' as follows:

[55] There are two widely used macro-design principles found in casino design: (1) the 'playground design', and (2) the Friedman-style 'gaming design' (Griffiths, 2009). The first type aims to create relaxing environments in which patrons are "more likely to spend their time and money" (Griffiths, 2009, p. 23). The second type uses design elements that focus player attention on the activity of gambling. This chapter addresses type (2) designs, since they are more overtly hostile to agents' interests.

> It's like being in the eye of a storm, is how I'd describe it. Your vision is clear on the machine in front of you but the whole world is spinning around you, and you can't really hear anything. You aren't really there—you're with the machine and that's all you're with. (Mollie, as cited in Schüll, 2012, p. 2)

The techniques employed by casinos to facilitate 'the zone' have been described as "entrapping mechanisms" that "exploit the cognitive expectations of new gamblers such that they persevere at the interaction to a point where the self-maximizing aim of winning turns into the self-liquidating aim of the zone" (Horbay, as cited in Schüll, 2012, p. 97). Here is the first of many instances where the interests of certain players and the house diverge. A player may have the local interest of maintaining a 'zone' state but the overall interest of keeping to a budget limit. The casino's aim of maximising profits—by increasing time on device—exploits the conflicted player. The scaffolding that was previously aiding affective self-management then becomes hostile. As Schüll (2012) explains:

> The affective appeals of casino design come palpably to the fore when they conflict with the conscious intentions of patrons— as in the case of gambling addicts attempting to resist the pull of machine play. (p. 49)

Furthermore, as we will soon see, casinos do not benefit by chance, they engineer these scaffolds—which puts them causally 'in the loop' of player exploitation (Friedman, 2000).

This section will begin by analysing shallow hostile scaffolding in casino interior design and ambience (Friedman, 2000; Schüll, 2012). I then provide two deep hostile cases: gambling machines that utilize player tracking and virtual reel mapping (Schüll, 2012). The latter examples demonstrate that deep hostile scaffolding is not only theoretical but also a real phenomenon.

## 3.3.1. Shallow Hostile Scaffolding in Vegas

### *Casino Interior Design*

> The job of casino layout is to suspend walking patrons in a suggestible, affectively permeable state that renders them susceptible to environmental triggers, which are then supplied. (Schüll, 2012, p. 46)

When a person enters a casino, they may have a variety of tasks other than gambling. These may include dining at restaurants, shopping, finding hotel lobbies or locating bathrooms and exits. As the above quote indicates, however, casino interiors are often heavily scaffolded with the aim of enticing people to gamble. Designers, such as Bill Friedman (2000), argue that "instead of turning attention away from machines, every aspect of the environment should work to turn attention toward machines, and keep it focused there" (Schüll, 2012, p. 40). For the customer with the explicit task of gambling, these interior scaffolds are not hostile: the scaffolding allows agents to easily locate gambling areas. Hostility occurs when the environment overrides an agent's previously defined (non-gambling) tasks and goals. The customer attempting to find a restaurant may instead be pulled deeper into the gambling area and anchored to a machine. In this section, I discuss some of the scaffolds that influence agents' movement and attention and show how they meet the criteria for shallow hostile scaffolding. My analysis includes 'equipment immediacy', maze-like floor layouts and the implementation of curving passages (Friedman, 2000; Schüll, 2012).

Friedman (2000) uses a number design techniques that all act in tandem to scaffold and direct players. One design element he discusses is 'equipment immediacy'—where customers are exposed to gambling activity as soon as they enter casinos (Friedman, 2000, pp. 43-44). This acts as a shallow cue for patrons and may particularly influence those with predispositions for gambling behaviour (Friedman, 2000, p. 147). Additionally, Friedman (2000) argues that the surroundings of gambling machines should employ 'The Law of Space Elimination'. This "dictates that designers 'constrict' space to create protected sanctuaries for play" (Schüll, 2012, p. 43). Coffers, hoods, canopies and soffits are all used to enclose what are actually vast spaces. While these elements influence affectivity—providing customers with a sense of security and privacy—they also make it difficult for "visitors to determine where they are and where they want to go" (Friedman, 2000, p. 79). One way to solve this navigational issue is to implement a maze-like floor design (with narrow isles) to direct movement. Friedman (2000) notes how mazes can focus the attention of their occupants, while simultaneously confusing them.[56] In this way, the environment steers player behaviour "in accordance with the extractive aims of the larger operation"—i.e., to get customers to gamble (Schüll, 2012, p. 40). As Friedman (2000) details:

> A maze layout rivets visitors' attention on the equipment immediately ahead. The slot faces at the ends of the short, narrow aisles are thrust right at them. The convoluted, dead-ended pathways force walkers to focus on the machines as they approach to avoid

[56] This can make it difficult for agents to navigate when they have non-gambling tasks.

bumping into them. If a visitor has a propensity to gamble, the maze layout will evoke it. (p. 147)

By implementing mazes, gambling equipment is not only thrust upon players as they enter casinos, but also at multiple points as they proceed through the interiors. Schüll (2012) explains that, despite a maze's usual associations with feelings of disorientation, "Friedman's maze shrinks and structures space in such a way as to orient patrons along a certain course" (p. 44). Due to the restriction of space, visitors can only "see a short distance ahead to the items directly in front of them" (Friedman, 2000, p. 63). This then "induces [them] to continue farther and deeper into the gambling equipment" (Friedman, 2000, p. 64).

The final interior design element I discuss is the use of gradual curving passages. These exterior entryways and interior passages smooth out "any edges or angles that might cause [walkers] to pause, shift, or reflect on their movement" (Schüll, 2012, p. 46). They also narrow gradually to ensure that patrons do not realise when they have transitioned into play areas (Schüll, 2012, p. 46). The effectiveness of gambling cues can be enhanced further when these passages replace features (such as sharp turns) that may cause agents to notice how they are being directed (Schüll, 2012). By rendering patrons 'unreflective', these passages act as a type of affective scaffolding. Altogether, these techniques scaffold casino interiors by directing movement, ensuring agents remain unreflective, all while supplying triggers that prompt an "emotional passion to gamble" (Friedman, 2000, p. 81).

As previously noted, when a customer enters a casino with the goal of gambling, these scaffolds are not hostile—they aid in directing the agent's attention towards their goal. One way that hostility can occur is when these scaffolds trigger the urge to gamble in those with non-gambling tasks. In a hostile scenario, we can imagine a casino customer (agent 1) moving through a casino with the task of engaging in non-gambling activity—i.e., eating at a restaurant or locating other facilities (such as bathrooms and exits). The interior design is scaffolded in multiple ways and allows casinos to: (1) make gambling equipment immediately visible to agent 1; (2) use maze-like layouts to funnel movement farther into the gambling area, ensuring more exposure to gambling cues; and (3) utilize smooth and gradual passages to keep agent 1 in an unreflective state (making them more susceptible to cues). These scaffolds work in tandem to change the cognitive demands of agent 1's navigational task. By increasing the amount (and potency) of gambling cues, the other tasks become more difficult to perform—especially if agent 1 is predisposed towards gambling. If agent 1's previous non-gambling task is overridden, then agent 2 (the casino) benefits financially by undermining agent 1's overall interests. Furthermore, agent 2 is causally 'in the loop' since these scaffolds are designed and implemented with the stated purpose of guiding behaviour towards gambling activity (Friedman, 2000, p. 147). At this point, it should be noted that deliberate intent is not a pre-requisite for hostility, since that would discount many of the previously discussed animal cases, but it does make salient that casinos benefit in ways that are causally relevant.[57] So, because agent 2 benefits through the subversion of agent 1's interests, and is causally 'in the loop', the scaffolding is hostile. Lastly, these interior scaffolds are shallow because they provide agent 1 with visual and directional cues.

The impact of such techniques can also be demonstrated through real-world accounts of agents attempting to resist these environmental pulls. Schüll (2012) recounts the experience of Todd, a man attending a Gamblers Anonymous meeting, as he walked through a casino in order to meet friends at a restaurant. In this instance, we can specify his task as locating the restaurant (while avoiding gambling activity). As he attempted this, he noted that the "architectural and atmospheric features, working in concert with its gambling equipment, triggered in him a powerful psychological and physiological reaction" (Schüll, 2012, p. 49). Despite Todd's previous decision to abstain from gambling, "his 'nervous system' (to use his term) was thrown for a loop" when exposed to the environmental stimuli present in the casino (Schüll 2012, p. 50). He repeatedly described himself as "lost", as well as the difficulty of averting his gaze from the direction of the machines (Schüll, 2012, p. 50). He also described how his eyes "found all the machines [he] liked to play" and how "[he] knew exactly where they were" despite never having entered that casino before (Todd, as cited in Schüll 2012, p. 50). Even though Todd did not gamble in this instance, it was highly likely that he could have.

Assuming that Todd (or a similar agent) did gamble, we can again summarize this case by using our hostile scaffolding table:

| | |
|---|---|
| Agent 1 (Who suffers?): | Todd |
| Agent 2 (Who benefits?): | The casino. |
| The task: | Moving through a casino with the task of locating a restaurant. |
| The external structure(s) (ES): | The interior design of the casino (i.e., the use of equipment immediacy, the maze-like layouts and passage design). |
| Does ES transform the cognitive demands of a task? | Yes. ES exposes agent 1 to gambling activity while making non-gambling navigation more difficult. |
| Does ES harm agent 1? | Yes. It harms agent 1 when the urge to gamble overrides previously defined non-gambling goals. |

[57] We will see this 'intention' in many of the upcoming cases.

| | |
|---|---|
| Does ES benefit agent 2? | Yes. The casino financially benefits when customers gamble. |
| Is Agent 2 causally 'in the loop'? | Yes. The casino work with designers and implement ES with the stated purpose of guiding customers behaviour towards gambling productivity. |
| Shallow or deep? | Shallow. The scaffolding provides agent 1 with cues: various visual stimuli and directional cues. |

For Todd, the interior scaffolds of the casino are hostile, since they attempted to undermine his interests while serving the financial interests of the casino. His account also demonstrates the difficulty of navigating a casino environment while attempting to avoid gambling activity. These navigational challenges can be highlighted further by Schüll's (2012) account of paramedics attempting to locate and treat heart attack victims within casinos (pp. 30-33). In an interview, one paramedic describes the confusing process of navigating a casino's interior as follows: "It all looks the same—you go up and down elevators, there are no direct routes, the carpets lead you around and around, you lose your sense of direction" (Schüll, 2012, p. 30). Due to the Friedman-style casino's focus on machines and the maze-like elements, the interior uses shallow hostile scaffolding to increase the cognitive demands of non-gambling navigational tasks in ways that serve the casino's interests (and goals) by promoting gambling activity.

### *Casino Ambience & Atmosphere*

Ambience and atmosphere also play large roles in scaffolding player behaviour and can make agents more susceptible to cues, as well as enhance certain aspects of the gambling experience (Schüll, 2012). By balancing lighting, colour, sound, temperature and smell, casinos are able to "elicit an emotional or physiological reaction from customers"—which can increase the likelihood of immersing agents in 'the zone' (Friedman, 2000, p. 84). For the gambling customer with unlimited time and budget, this type of scaffolding is not hostile—since it aids in perpetuating long play sessions. But most gamblers will need to manage their time and budget. Casino ambience becomes hostile scaffolding when it contributes to the agent exceeding these limits.

Let us begin with lighting and noise, which are carefully balanced to limit player distraction. Friedman (2000) argues that the presence of overly bright lights may result in customers ending gambling sessions sooner or decrease the likelihood of players returning—even if they are unaware of what elicited those feelings. Choosing dimmer lights, as well as reducing vivid or conflicting colours, therefore aids in placing less sensory strain on patrons, making them more likely to continue gambling.[58] Similarly, it is argued that noise should also be balanced to reduce strain—i.e., music should not be too soft or heavy (Friedman, 2000, p. 135). Most importantly, Friedman (2000) notes that noise should not distract players by being deflected around them.[59] The stated aim of balancing environmental ambience is to "powerfully modulate patrons' 'experiential affect', not only helping to usher them to machines, but to immerse them in the zone, and keep them there" (Schüll, 2012, p. 46). Just as customers may be consciously unaware of the elements that push them away from gambling, they may also be unaware of how balanced ambient elements compel them to gamble more. Schüll (2012) argues that these "ambient strategies treat affect not as something passive or static, but as an active and dynamic capacity that can be harnessed and guided in lucrative directions" (p. 46). Casinos are, therefore, causally 'in the loop' since these scaffolds are adjusted with the explicit aim of reducing customer sensory stimulation and distraction, while promoting 'zone' immersion (Friedman, 2000). The less likely a customer is to become self-reflective or distracted, the less likely they are to cease playing.

We can identify two more atmospheric elements that have strong affective influences on gamblers: music and smell. In §1.3, I discussed how music could aid in regulating an agent's affective states. Recall how a workout playlist could provide extra motivation when exercising, or how a study playlist may assist with concentration. Casino operators also use music as affective scaffolding. Digigram is a company that provides casinos with music that guides customer behaviour—which they refer to as "functional music" (Schüll, 2012, p. 48). They cite studies that indicate how music can influence walking speed, as well as the time and money spent in retail spaces (Schüll, 2012, p. 48). In order influence certain customers, Digigram provides casinos with systems that change background music depending on the time of day and the profile of clientele. The company recommends music that is "slow or mild in the middle of the day for one group of customers" and then "build[ing] up the tempo throughout the day when there's a high occupancy of customers" (Manager for Digigram, as cited in Holtmann, 2004, p. 30). They emphasise: "You [the casino operators] have control of the ambience" (Manager for Digigram, as cited in Holtmann, 2004, p. 30). DMX music, another casino sound supplier, also describes their objective as "assist[ing] [casinos] in stimulating their customers to respond to their

[58] It could be argued that casinos not featuring wall clocks or ceilings that connect with natural light can further distort a player's sense of time, but Griffiths (2009) argues that "there is little empirical research on the effect that this has on players" (p. 26). It is for this reason that the absence of clocks and natural light will not be discussed here.

[59] Friedman (2000) observed that players show signs of fatigue, tension, and distress when heavily deflective surfaces are present (p. 136).

environments" (Holtmann, 2004, p. 30). These music systems work together with balanced lighting and noise levels to make customers more likely to start gambling, as well as encouraging "the flowing suspension of the zone state by supplying patrons' perceptual systems with a subtle, even-keeled stream of sensory input" (Schüll, 2012, pp. 48-49). In this way, music can act as a "behavioral modulator" that operates in the background (Schüll, 2012, p. 48). Schüll notes how "known tunes and slow passages that do not vary in volume and rhythm work well to orchestrate consumer action while remaining below the threshold of consciousness" (Schüll, 2012, p. 48). In a similar manner to gradually curving passages, casinos aim to keep customers in cognitively unreflective and suggestible states. Conversely, varied music may disrupt gambling by "restor[ing] . . . [the gambler's] cognitive state to where [they] can make rational decisions" (Finlay, as cited in Thompson, 2009, n.p.).

Lastly, smell can also influence a player's affective states and increase time on device. Schüll discusses a paper by Hirsch (1995) which found a significant increase in profit (45%) at slot-machines that had "been subtly treated with a certain pleasing odor" (Schüll, 2012, p. 47). Hirsh (1995) speculated that the emitted odours created an "affective congruence with the situational context," promoting longer play sessions (p. 593). Additionally, when an odour is "matched to a certain environment", it can "precipitate actions" by eliciting conditional responses (Hirsh, 1995, p. 593). Pleasant odours therefore serve as another cue to keep customers playing for longer periods of time.

When taken together, lighting, noise, music and smell are all controlled by casino operators to scaffold play spaces in ways that make agents unreflective, susceptible to triggers and more immersed in 'the zone'. Let us review how casino ambience meets the criteria for shallow hostile scaffolding. Agent 1 (the gambler) may have the task of gambling within a time limit. The casino ambience and atmosphere (consisting of balanced lighting, sound, music and certain aromas) scaffolds agent 1's cognition by increasing the duration of 'zone-like' play. The scaffolding makes it easier for agent 1 to focus on the task of gambling but also makes them unreflective about other factors, such as the duration of their play sessions. Agent 1 may become so fixated by play that they exceed their previously allotted time—which can be financially detrimental. The longer customers play, the more likely they are to spend. Agent 2 (the casino) then benefits when a player extends their time on device. This atmospheric scaffolding is also hostile because the casino is causally 'in the loop': they balance and control lighting, acoustics, music and aromas in order to elicit player focus. Finally, the scaffolding is shallow because the controlled elements cue agents to respond in certain ways.

The case for ambience as hostile scaffolding is summarised in the table below:

| | |
|---|---|
| Agent 1 (Who suffers?): | The gambler. |
| Agent 2 (Who benefits?): | The casino. |
| The task: | To gamble within the boundaries of a time limit. |
| The external structure(s) (ES): | Casino ambience: balanced lighting, sound, music and certain aromas. |
| Does ES transform the cognitive demands of a task? | Yes. ES makes it easier to focus on the task of gambling while increasing player absorption. |
| Does ES harm agent 1? | Yes. Agent 1 may become so fixated on play that they lose track of time. The agent that plays more, spends more. |
| Does ES benefit agent 2? | Yes. The casino financially benefits by increasing 'zone' states, which extend time on device. |
| Is Agent 2 causally 'in the loop'? | Yes. The casino deliberately chooses certain lighting settings, acoustic environments, songs and aromas to elicit player focus. |
| Shallow or deep? | Shallow. All of these scaffolds cue agent 1 to respond in certain ways. |

In this way, casino ambience and atmosphere meet the criteria for shallow hostile scaffolding. It has also been shown that, when absorption in 'the zone' is heightened, players become much more oblivious to the world around them. Schüll (2012) notes: "As their [player] absorption increases over a session of play, the more impervious it becomes to potential distraction from disequilibrated or otherwise disruptive ambient elements" (p. 49). When reviewing video footage of heart attack victims, Schüll (2012) describes the most concerning aspect (aside from the actual heart attack) as the reaction from other players. They continued gambling, "despite the unconscious man lying quite literally at their feet" (Schüll, 2012, p. 33). When a player enters a 'zone' state, their entire focus is fixated by the gambling machine's screen. Together with the machines themselves, casino ambience is adjusted with the specific aim of heightening that initial 'zone' entry point and absorption. The resulting fixation also means that a players can lose track of previously defined goals, such as time and budget limits.

The above cases also raise questions regarding agent 'trust'. In §1.6, I drew from Sterelny (2010) to discuss how trust between an agent and scaffolding was often required before the scaffolding could be used effectively. But these hostile casino cases are able to establish trust even though agents' interests are undermined. How is does this happen? Part of the answer can be found in how the scaffolding is shared by many agents—which is particularly noteworthy since Sterelny (2010) highlighted that the sharing of resources made hostility more unlikely. In the case of a fake subway maps, Sterelny (2010) noted that it would be difficult to manipulate an agent in specific ways (since there are too many variables) and that the shared nature of these spaces was a type of extra defence (due to multiple reliable agents being involved). But hostile scaffolding in casinos gives us reason to take the possibility of hostility in shared spaces more seriously. What is not considered by Sterelny (2010) is that some shared spaces still allow for large degrees of control and manipulation. They are set up in ways that precisely influence agents in ways that designers (and beneficiaries) intend. Furthermore, the fact that many agents use these spaces could improve agent trust in otherwise hostile environments. As demonstrated with casino design and ambience, casinos can take advantage of the shared spaces to profit in hostile ways (Friedman, 2000; Schüll, 2012).

### 3.3.2. Deep Hostile Scaffolding in Vegas

So far, I have analysed three instances of shallow hostile scaffolding. Not only do these cases highlight the presence of hostile scaffolding in everyday life, but they also demonstrate the application of the criteria developed in Chapter 2. I will now shift my focus to deep hostile cases and show how they meet all the same criteria but, instead of involving cues, the scaffolding supports the offloading of significant cognitive processing. I start with two more casino examples and detail how gambling machines use statistical tracking and virtual reel mapping as deep hostile scaffolding (Schüll, 2012).

*Player Tracking & Interventions*

Cases of deep benign scaffolding that use tracking technology can be easily found. Today's smartwatches come equipped with trackers that calculate users' stress scores. They can monitor a variety of factors to determine the score such as heartrate, blood oxygen levels and electrodermal activity. Once this data has been collected, the devices analyse and display the user's daily stress assessment. The user may then be provided with strategies to improve their score, which can include breathing exercises and mindfulness techniques. This is an instance of deep benign scaffolding. The smartwatch collects data and performs external calculations to facilitate a portion of the cognitive work involved in determining stress levels. Once the information is processed, it then loops back to agents and allows them to respond appropriately. Casinos, on the other hand, use statistical tracking as deep hostile scaffolding. The accumulated information gathered over a player's entire gambling history is used to influence their affective states in ways that encourage more gambling activity (Schüll, 2012). For example, a player may have the task of determining when to cease gambling activity, with most players having a particular loss threshold. After losing a certain amount, players are likely to walk away. Casinos use tracking systems to predict player 'pain points' and intervene by offsetting the negative feelings associated with continued losses (Schüll, 2012). I argue that such cases are hostile variations of the smartwatch example and are, therefore, genuinely deep.

Casinos first implemented player tracking in 1985 and mimicked the reward systems of airlines and credit card companies (Schüll, 2012, p. 144). Punch cards were notched each time players hit jackpots, which could then then "redeem the cards for meals and other rewards" after a certain number of notches was reached (Schüll, 2012, p. 144). This provided casinos with data while also affectively motivating players—the greater the number of notches, the more likely they were to play again. Modern player tracking has become more advanced. Gambling machines now allow casinos to track the value of every bet, a player's win and lose rate, the rate of button pressing, the timing of breaks, as well as food and drink purchases (Schüll, 2012, p. 144). A player's data is no longer captured over one sitting but their whole history and habitual modes of play can be recorded. In 2005, Harrah formulated a way to "optimize" player value (Schull, 2012, p. 154). This system calculated how much a player could lose before exhibiting negative responses—such as ceasing gambling activity altogether. A 'pain point' was then created based on each individual players' observed thresholds. Once this point was reached, the casino attempted to intervene by dispatching a 'Luck Ambassador'—who would offer positive reinforcement by giving out vouchers or show tickets (Schüll, 2012). The intended result was to mitigate the negative emotions associated with continued losses—meaning players would continue gambling or, at least, take shorter breaks between sessions (Schüll, 2012). Machines that utilise these systems can be said to be deep hostile scaffolds. They are deep because they use previous and real-time player data to influence agents' affective states, much like the smartwatch. But unlike smartwatches, the main aim of casinos is to manipulate a player's affective state to prolong gambling activity, despite some players reaching their loss thresholds. Increasing time on device can benefit the house at the expense of players—with players potentially exceeding their budget or predefined goals (such as quitting after X losses). However, the system ultimately proved unsuccessful. Players became irritated by their games being constantly interrupted, which also impacted their ability to enter 'zone' states (Schüll, 2012). When entering 'the zone' players seek privacy and focus (Friedman, 2000). Interruptions thereby replace the intended positive reinforcement with frustration.[60]

Even though the Luck Ambassador system failed, less disruptive interventions would later be implemented with greater success. International Gaming Technology's marketing system, Experience Management, has been described as giving casino operators the tool set to "talk to and market to a consumer while he's actually consuming the product" (Schüll, 2012, p. 162). Schüll explains that this system functions "like a fully digital version of the Luck Ambassadors system" (Schüll, 2012, p. 162). Unlike the previously interruptive Luck Ambassadors, these interventions motivated player behaviour towards continued play by distributing immediate incentives on screens. Casinos can "instantly credit players with rewards when they reach their personal 'pain points'" by generating and displaying an offer at the exact moment a pain point is reached (Schüll, 2012, p. 162). We now have our first successful, real-world case of deep hostile scaffolding. Here, agent 1 is the gambler and agent 2 is the casino. We can identify agent 1's task as ceasing gambling activity after X losses, or after they have reached their budget limit.[61] The scaffolding is the gambling machine that tracks players and is linked to an instantaneous marketing system (such as Experience Management). The machine is deep scaffolding because it performs significant processing in order to determine agent 1's pain point. It then loops back to agent 1 (influencing their affective states) by providing rewards (such as credits). When agent 1 gambles past their previously defined limit, it can be harmful in ways that benefit agent 2—i.e., when time on device is extended, the casino extracts more money. Furthermore, agent 2 is causally 'in the loop' because casinos intentionally calculate player pain points and intervene in order to influence player affective states—financially profiting as a result.

Let us place machines that use tracking and marketing systems into our hostile scaffolding table:

| | |
|---|---|
| Agent 1 (Who suffers): | The gambler. |
| Agent 2 (Who benefits?): | The casino. |
| The task: | To determine when to cease gambling activity after losing a certain amount. |
| The external structure(s) (ES): | The gambling machine with tracking and linked to an instantaneous marketing system. |
| Does ES transform the cognitive demands of a task? | Yes. The machine and marketing system offset negative emotions by providing credit |

[60] A consultant at Harrah explained: "One woman got so frustrated that she put her own five dollars in the machine, to get them to go away and let her play" (Schüll, 2012, p. 169).
[61] Their primary task may be to gamble but, in this case, the agent also sets an additional task of ceasing gambling activity after reaching a predetermined limit.

| | |
|---|---|
| | rewards to players as they reach their calculated pain points. |
| Does ES harm agent 1? | Yes. The customer is incentivised to continue gambling despite reaching their pain point. This may result in an agent exceeding their budget. |
| Does ES benefit agent 2? | Yes. When a player extends their time on device, more money can be extracted. |
| Is agent 2 causally 'in the loop'? | Yes. The casino calculates player pain points with the intension of intervening and extending play. |
| Shallow or deep? | Deep. The technology utilised by the machines prolongs play by gathering and processing player data to manipulate player affective states. |

These machines tick the same hostile scaffolding boxes as the Luck Ambassador cases but are actually successful in promoting continued play, due to their non-disruptive interventions. They are deep hostile scaffolds that, in similar ways to smartwatches, monitor behaviour, and influence affective states. Unlike the stress monitoring smartwatch, however, casinos gain by shifting affectivity in ways that undermine players and benefit the casino.

Returning to Sterelny's (2010) dimensions, the shared nature of gambling machines seems to imply that they are interchangeable and not individualised. However, the machines in the above deep case are able to be both. Machines that utilise player tracking systems adapt to player input and dispatch rewards depending on each player's observed pain threshold. Additionally, some machines utilise behavioural intelligence software to adapt to other player preferences—such as the skipping of animations to allow for faster play (Schüll, 2012, p. 169). These have similar effects to Sterelny's (2010) example of a set highly modified set of chef's knives (p. 475). When machines adapt to players, an "enhanced sense of interaction between player and machine is created" (Giacalone, 1996). As we discussed in §1.6, individualising a scaffold may improve complementarity and entrench it deeper into to an agent's cognitive system. Machines that adapt to players can therefore be deep scaffolds that also meet the criteria of being individualised.

### *Slot Machines & Virtual Reel Mapping*

Early slot machines displayed relatively honest odds and results. They were purely mechanical devices, operated by gears and springs. The Liberty Bell machines were the prototypes for modern reel-spinning machines and consisted of three spring-loaded reels, depicting five symbols: bells, horseshoes, diamonds, spades and hearts (Schüll, 2012). As Schüll (2012) explains:

> When three bells matched up across the central payline, a prize of ten nickels was rendered. The reels rotated around a supporting metal shaft connected to a handle mechanism and a braking system, and a timing bar stopped the reels one at a time from left to right to create suspense. (p. 80)

Later machines would increase the number of symbols (possible 'stops'), reducing the odds of winning, and would reach 22 by 1970 (Schüll, 2012). Lowering the winning odds meant that casinos could offer higher pay-outs while remaining profitable. Additionally, affectively motivating elements were added. For example, the viewing window was expanded so "players could see rows of symbols above and below the payline, increasing the likelihood that they would experience a 'near miss'" (Schüll, 2012, p. 80). A near miss—which is "the sensation of nearly having won produced by the sight of winning symbols adjacent to the payline"—encourages more attempts (Schüll, 2012, p. 81). As Skinner notes: "Almost hitting the jackpot increases the probability that the individual will play the machine, although this reinforcer costs the owner of the device nothing" (Skinner, 1953, p. 397). Yet despite these developments, most of the older machines could be classified as benign scaffolding. They were external structures that changed the cognitive demands of determining gambling results by displaying easily interpretable symbols. The manner in which symbols were displayed served as reliable, consistent and accurate representations of the actual outcomes and odds of winning. In this way, agents could also use the machines to accurately decern the odds of hitting a jackpot. Furthermore, if near misses did occur, they were genuine. As such, the results displayed by these machines were not misrepresentations.

As machines continued to develop, they began incorporating electromechanical elements (such as motors, circuit boards and digital microprocessors) in the process of reel spinning (Schüll, 2012, p. 81). When random numbers were generated by microprocessors, they were translated to stops on the microprocessor's 'virtual reels' and "communicated to the correlating positions on the actual, physical reels" (Schüll, 2012, p. 82). Machines became overtly hostile to players' interests with the development of 'virtual reel mapping' (patented in the 1984 by Inge Telnaes). Through this process, the number of non-paying virtual reels are disproportionately mapped to the number of actual stops—making the odds of winning appear higher than they actually are (Schüll 2012, p. 86; Telnaes, 1984). Once the random number generator (RNG) value is produced, a secondary mapping process assigns more stops to "low-paying or nonpaying blank positions on the actual, physical reel than to winning positions" (Schüll, 2012, p. 87). With this type of mapping, casinos could offer even greater jackpots by lengthening the odds, while still maintaining the appearance of 22 reels. Contrasted with older reels, these virtual reels could be "configured to accommodate as many stops as designers like, sometimes hundreds" (Schüll, 2012, p. 87). The figure below offers a visual representation of this process:

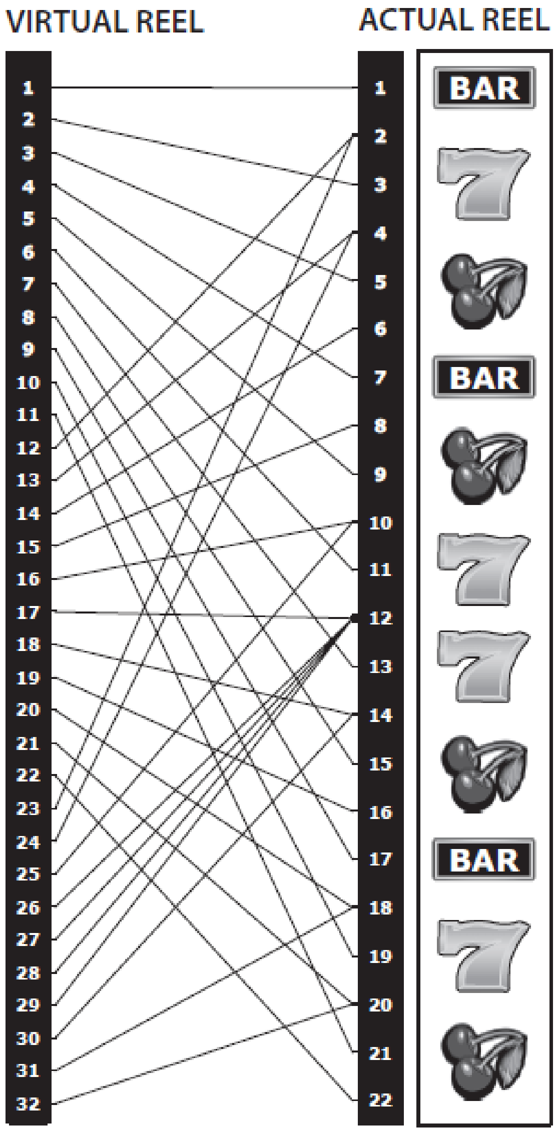


Figure 1. Educational illustration of virtual reel mapping by Game Planit, Inc. (From Schüll, 2012, p. 88)

Schüll explains:

> Although each symbol that players see seems to have an equal chance of hitting, in fact each does not; the actual reel merely communicates the mapping decisions of its much-expanded virtual counterpart. Telnaes wrote candidly of his intent to distort player perception: "It is important to make a machine that is perceived to present greater chances of payoff than it actually has." (Telnaes, 1984, as cited in Schüll, 2012, pp. 89-90)

Telnaes' (1984) admission of distorting player perception places casinos causally 'in the cognitive loop' when cognitive distortion occurs. Early attempts at increasing the number of stops saw designers use extra or larger reels to accommodate more symbols (Schüll, 2012). However, because this could be perceived by players as accurate representations of the increased odds, these machines were often avoided (Schüll, 2012). Casinos are aware that these dishonest forms of reel mapping increase profitability and the techniques were explicitly designed to distort player perception.

In addition to misrepresenting the odds, virtual reel mapping also affectively motivates players when coupled with clustering. 'Clustering' refers to a technique that allows designers to "map a disproportionate number of virtual reel stops to blanks directly adjacent to winning symbols on the physical reels" (Schüll, 2012, p. 92). Winning symbols can then appear above and below the blanks of the pay line, "far more often than by chance alone" and allows for the frequency of near misses to increase (Schüll, 2012, p. 92). We have already established that near misses are affectively motivating and, through clustering, casinos can better recontextualise actual losses as possible wins. Players are then more likely to continue playing, which increases the amount of money that casinos can extract.

So, how do slot machines that use virtual reel mapping and clustering meet the criteria for deep hostile scaffolding? Again, the gambler is agent 1 and the casino is agent 2. We can set two tasks for agent 1: (1) choose a gambling machine after decerning the odd of winning, and (2) decide whether or not to spin again after losing. For task 1, the slot machine uses virtual reel mapping to distort player perception by making machines appear more likely to hit jackpots than they actually are. Agent 1 therefore decides to gamble on a machine that they would have otherwise avoided, which can be financially detrimental. For task 2, clustering creates near misses that affectively motivate agent 1 to spin again. If agent 1 has a previously determined budget limit, they are harmed when near misses coerce them to exceed this limit. Agent 2 is causally 'in the loop' because the stated aim of virtual reel mapping is to "make a machine that is perceived to present greater chances of payoff than it actually has" (Telnaes, 1984). Similarly, clustering intentionally creates more near misses to motivate additional spins. When agent 1 chooses a machine with odds lower than what is perceived, and they are then motivated to perform additional spins, agent 2 financially profits. Finally, machines that utilise these techniques are deep because clustering and virtual reel mapping facilitate significant cognitive processing. Players learn that the number of symbols indicates probability, and the ways that the symbols spin and stop can be used to assess success or failure. The assessment of probable success (the virtually mapped symbols) and actual success (the near misses resulting from clustering) are thereby transformed in deep ways by the scaffolding.

In summary:

| | |
|---|---|
| Agent 1 (Who suffers?): | The gambler. |
| Agent 2 (Who benefits?): | The casino. |
| The task: | (1) Choosing a machine by discerning its winning odds.<br>(2) Decide whether or not to spin again. |
| The external structure(s) (ES): | The slot machine using virtual reel mapping and clustering. |
| Does ES transform the cognitive demands of a task? | Yes.<br>(1) ES uses symbols and virtual reel mapping to make the winning odds appear higher than they actually are.<br>(2) Clustering results in agent 1 deciding to spin again. |
| Does ES harm agent 1? | Yes.<br>(1) ES distorts perception and results in agent 1 gambling at machines with lower odds. This is manipulative and financially detrimental.<br>(2) Clustering creates affectively motivating 'near misses' that may result in agent 1 exceeding their monetary budget. |
| Does ES benefit agent 2? | Yes. The casino financially benefits when customers gamble at machines that misrepresent odds. 'Near misses' then encourage more time on device. |
| Is Agent 2 causally 'in the loop'? | Yes. The stated aim of these techniques is cognitive distortion and affective motivation. |
| Shallow or deep? | Deep. Significant cognitive processing is facilitated by ES algorithms and displays. |

It has been argued that machines that use these techniques "hoodwink the human perceptual system and encourage player persistence" (Harrigan, 2007, as described in Schüll, 2012, p. 90). By glancing at the above table, we can see that machines using virtual reel mapping and clustering techniques are deep hostile scaffolds: they facilitate significant processing, but intentionally distort player perception. They can lead to gamblers choosing unfavourable devices and motivate them to keep playing.

## 3.4. Deep Hostile Scaffolding & Gamification

My final example of deep hostile scaffolding considers how Twitter negatively gamifies agents' lives by shifting their communication goals to benefit the platform (Nguyen, 2021). Gamification can be defined as "the introduction of game-like elements into practical life" (Nguyen, 2020, p. 189). For example, Duolingo gamifies how we learn languages by giving us daily goals, rewards and leaderboards, and FitBits encourage, track and reward daily exercise goals. Nguyen (2020) argues that gamification is not negative when it increases our motivation by giving us simple and clearly defined goals (such as language learning). Issues start to arise when gamification overly simplifies an agent's values. Nguyen refers to this phenomenon as 'value capture'—where complex and nuanced values are overridden with simplified ones (Nguyen, 2020, p. 201). Platforms such as Twitter gamify communication in ways that override an agent's previously established communication goals in favour of higher 'scoring' posts (Nguyen, 2021). This is hostile when it undermines the interests of agents attempting to communicate while serving the interests of the platform—increasing platform usage and advertiser revenue.

Twitter not only provides a platform that allows a vast number of people to communicate, but it also "suggests specific goals for those interactions" (Nguyen, 2021, p. 1). Nguyen (2021) notes how Twitter gamifies communication through its implementation of scoring systems—with Likes, Retweets and Follower numbers. Some of our 'ordinary' goals for communicating with others may be to "pursue truth and understanding, or to promote empathy for one another" (Nguyen, 2021, p. 11). But Twitter replaces these nuanced goals with simpler ones, such as "maximizing one's Likes, Retweets, and Follower counts" (Nguyen, 2021, p. 11). A tweet with a higher number of Likes and Retweets is shared more often and seems to suggest 'better' communication content than those with lower scores. Much like Slaby's (2016) idea of 'mind invasion', Nguyen (2021) warns that, if we subscribe to and internalize Twitters goals and value systems, we may negatively shift our own communication goals. Tweets that use over-inflammatory language, or that aim to induce strong moral emotions (like outrage), may do better than tweets aiming to be factual (Nguyen, 2021). Additional ways Twitter shifts communication include: a binary scoring system that only gives positive or negative results with little room for nuance; and how the act of liking a tweet is normally done on first impression bases, rather than after serious thought or discussion (Nguyen, 2021, pp. 11-14). Furthermore, Nguyen (2021) suggests that Twitter's approach to gamification was not designed with the interest of supporting "the plurality of communicative values" (p. 14). Instead, we have more reason to suspect that "its design features were heavily driven by an interest in increasing user engagement for the sake of profit" (Nguyen, 2021, p. 14). The more users engage with the platform, through positive or negative content, the more Twitter profits.

With this in mind, let us examine how Twitter's scoring system qualifies as deep hostile scaffolding. Consider a mobile phone with the Twitter app installed. Agent 1 is the Twitter user and agent 2 is Twitter. Agent 1 has the task of deciding on the content of a tweet regarding a recent event and observes how the platform widely displays inflammatory tweets. The platform distributes these tweets to users who then feed the algorithm through Likes and Retweets, indicating appropriate communication content. Agent 1, who was previously neutral on the topic, then decides to write an outrageous tweet to maximize their own score. Over time, agent 1's 'good faith' communication goals are overridden by negative ones (such as using language that elicits outrage or contains falsities). As this content becomes increasingly viral, agent 1's communication goals shift and agent 2 benefits from the increased user engagement (profiting from advertiser revenue). Agent 2 is causally 'in the loop' since the scoring systems and algorithms were designed to promote content that maximizes user engagement. This also scaffolds communication by shaping how agents choose to engage with others. Finally, this is a deep process because the platform supports the cognitive work (involved in agent 1 choosing how to communication) through a scoring system. This system performs calculations to score tweets and then distributes content back to users, suggesting certain communication goals. When this changes agents' previously defined goals, it is hostile scaffolding—as agent 1's values are overridden in order to benefit agent 2.

We can sum up our Twitter example with a final hostile scaffolding table:

| | |
|---|---|
| Agent 1 (Who suffers?): | The Twitter user. |
| Agent 2 (Who benefits?): | Twitter. |
| The task: | Deciding on the content of a tweet. |
| The external structure(s) (ES): | A device connected to the platform |
| Does ES transform the cognitive demands of a task? | Yes. Twitter rates content based on Likes and Retweets (which is fed back to agents) and this suggests agent 1's communication goals. |
| Does ES harm agent 1? | Yes. When this process involves 'value capture', agent 1's previously good |

| | |
|---|---|
| | communication goals are overridden by negative ones. |
| Does ES benefit agent 2? | Yes. Twitter maximises profits the more users engage with content (positive or negative). |
| Is Agent 2 causally 'in the loop'? | Yes. The platform was designed with the interest of maximizing user engagement. |
| Shallow or deep? | Deep. The platform performs calculations to score tweets and then distributes content back to users, suggesting certain communication goals. |

Twitter is therefore another instance of deep hostile scaffolding. Additionally, the ways in which scaffolding could gamify our lives in hostile ways can be explored further and Nguyen's (2020) other work could reveal similar cases (which will be discussed in §4.1).

## 3.5. Conclusion

After developing and laying out the case for hostile scaffolding in previous chapters, this chapter provided some much-needed examples—demonstrating that hostile scaffolding is real and relevant. I reviewed three additional shallow cases in the form of sunglasses (Viola, 2022), casino interior design and casino ambience (Friedman, 2000; Schüll, 2012). These use cues to undermine one agent while benefitting another. Most crucially, I then provided three deep cases to show that deep hostile scaffolding is not only theoretically but actual. My examples explored gambling machines that utilise player tracking technology, virtual reel mapping (Schüll, 2012), as well as Twitter's scoring system (Nguyen, 2021). By facilitating cognitive work, these scaffolds become extensions of agents' cognitive processes and are therefore deep cases. It should, however, be noted that my analysis of scaffolding in gambling is not extensive, and there remain additional cases to be addressed in the future. Hostile scaffolding may also extend into video poker and online gambling (Schüll, 2012). That said, I will not address those areas in this dissertation. This chapter's main aims were to provide additional hostile scaffolding cases and show that deep hostile scaffolding is possible. The next and final chapter will explore other future research avenues as well as conclude the dissertation.

# Chapter 4: Future Research & Conclusion

## 4.1. Future Research

Before concluding the dissertation, I will briefly review future avenues for hostile scaffolding research. These include the ethical implications of deep hostile scaffolding, scaffolding's role in forming addictive behaviour, additional work on gamification, 'racist scaffolding', and considering hostile scaffolding in developing technologies (such as virtual reality, dark patterns & AI).

### *The Ethical Consequences of Deep Hostile Scaffolding*

In §1.5, I argued that deep scaffolding is used in extended (hybrid) cognitive processes. Under the complementarity thesis, internal and external vehicles should be considered parts of a unified cognitive system when performing certain cognitive tasks[62] (Menary, 2006; Sutton, 2010). As technology develops and scaffolding becomes increasing coupled with our cognitive processes, more cases of deep hostile scaffolding could also arise. If the examples I provided in §3.3.2 and §3.4 are legitimate, and deep hostile scaffolding is real, then we should expand our legal framework to address additional ethical considerations.

One implication is already being considered: Can the intentional damage of cognitively integrated objects count as personal assault? (Carter & Palermos, 2016). In this instance, our cognitive abilities (and personhood) may become so tied to our reliance on certain artefacts that damaging them is equivalent to damaging one's person (Carter & Palermos, 2016, p. 5). Adam Carter & Orestis Palermos (2016) put forward the *Argument for Extended Assault* (AEA) as follows:

> (P1). Intentional harm to a part of a person which is responsible for her mental and other faculties constitutes personal assault. [Definition]
>
> (P2). Our mental faculties can be partly constituted by external artifacts, so long as these artifacts have been appropriately integrated into our overall cognitive system. [from HEC]

[62] Examples include using a calendar to determine availability.

(C). Therefore, having our integrated epistemic artifacts intentionally compromised plausibly qualifies as a case of personal assault [From P1 & P2] (p. 11)

The main crux of Carter & Palermos' (2016) argument is that, if the hypothesis of extended cognition (HEC) is true, then "our ethical and legal theorizing and practice should be prepared to include intentional damage of appropriately integrated gadgets within the category of personal assault" (p. 12). *Riley v. California* indicates that legal systems are already beginning to view integrated cognitive artefacts as more than just objects (Carter & Palermos, 2016). In this case, David Riley's conviction was overturned by the U.S. Supreme Court due his smartphone being searched without a warrant. Chief Justice, John Roberts, wrote that cell phones "are now such a pervasive and insistent part of daily life that the proverbial visitor from Mars might conclude they were an important feature of human anatomy" (Roberts, as cited in Carter & Palermos, 2016, p. 13). A cell phone should therefore not qualify as just another item in an arrested person's 'physical area'. Such cases indicate that AEA could plausibly be a consideration for legal systems in the future.

We can alter AEA to create a version that considers deep hostile scaffolding as well. This could include the ways in which integrated scaffolds manipulate cognitive processes in exploitative ways. A rough version of this argument could take the following form:

(P1) Deep scaffolds can be 'genuinely' cognitive parts of our mental faculties that perform tasks (following the complementarity principle).

(P2) The intentional manipulation of the deep scaffolds responsible for a person's cognitive faculties constitutes a form of cognitive exploitation when they harm agents (such as bringing about a harmful false belief.)

(P3) Bringing about a harmful false belief by using deep scaffolds is comparable to harmful instances of lying.

(C). Therefore, intentionally manipulating deep scaffolds in harmful ways should be subject to appropriate regulation and causally relevant agents should be held accountable for cognitive exploitation.

We could see this applied to gambling machines that utilise virtual reel mapping, since those techniques involve instances of harmful cognitive distortion and may be comparable to lying. Suppose a salesperson lies about their new vaping liquid—that it is safe and non-addictive—when it is actually more harmful and addictive than other products. If they lie and bring about a false belief that harms other agents, they are responsible for that false belief and may be subject to legal repercussions. If deep scaffolding were to deceive agents in comparably harmful ways, then (*prima facie*) the ('in the loop') beneficiaries are comparably responsible.

### *Addiction & Scaffolding*

When considering how agents become addicted, the role of engineered environments and scaffolding should not go unexamined. Don Ross (2020) notes how humans and other mammals both experience intoxication, yet the latter do not become addicted. Elephants and baboons can encounter fermented berries that lead to them "indulg[ing] in benders that they evidently find sufficiently rewarding" (Ross, 2020, p. 6). They are "at no risk of addiction, however, because they cannot cultivate sources of low-toxicity alcohol" (Ross, 2020, p. 6). Humans, on the other hand, are able to "engineer addictive environments" (Ross, 2020, p. 6). As discussed in Chapter 3, casinos use scaffolding to facilitate gambling activity which can result in problem cases. In order to mitigate responsibility, gambling vendors often state that such cases are "confined to a small minority of constitutionally predisposed or mentally disordered problem gamblers" (Abbott, 2006, p. 7); and that gambling machines are simply "the mechanism through which pre-existing psychological disturbances are expressed" (Blaszczynski, 2008, p. 7). But, contrary to this view, Schüll (2012) suggests that problem gambling results from the interplay between agents and objects, with some objects being "more likely than others to trigger or accelerate an addiction" (p. 20). How machines (and play areas) are scaffolded to "yield reinforcing rewards" should not be excluded from the discussions around gambling addiction (Ross, 2020, p. 3). The addictive effects on the brain include: (1) influencing the dopamine-based learning circuit to encourage further gambling; (2) teaching the ventral striatal circuit[63] to not settle on a mode of "genuine randomness" leading to further activity; and (3) cuing cravings (when not gambling) and rewards (when gambling) by directing focus (Ross, 2020, p. 3). Our analyses in Chapter 3 support the effects laid out by Ross (2020): experience management systems can influence the brain's reward systems; virtual reel mapping manipulates how odds are represented, as well as how results are displayed; and floor layouts can supply and cue gambling cravings. These scaffolds could therefore be viewed as fostering addictive behaviour. For Ross (2020), we should not look at our brain's learning systems as inherently flawed—since it did not evolve to selected against our "social capacity to manufacture and efficiently distribute flows of addictive products" (p. 7). Instead, we should examine how our engineered environments and products result in addition. If

[63] The circuit that influences decision making and reward-related behaviour (Ross, 2020, p. 3).

we agree with this, then the addictive properties of the hostile scaffolding present in §3.3 should be explored further.

### *Racist Scaffolding*

In §2.5, I discussed Liao and Huebner's (2020) oppressive things: material objects that could be racist when congruent with oppressive systems. We can apply this idea to other scaffolding cases as well. Recall how emergency exit and bathroom signs scaffold cognition by providing agents with cues that reduced the cognitive demands of locating certain facilities. The South African apartheid government also employed signage to scaffold cognition but did so in ways that were congruent with the racist systems of the time. For example, 'whites only' signs scaffolded bathrooms, beaches, benches and bus stops. These signs not only scaffold cognition by offering cues—suggesting who could use certain facilities—but they also meet Liao and Huebner's (2020) criteria for oppressive things. They are: (1) "biased in the same direction as other manifestations of an oppressive system", (2) are "causally embedded in the respective oppressive system", and (3) have casual connections that are bi-directional due to being products of the oppressive systems while also "guid[ing] and constrain[ing] racist psychological processes and racist social structures" (Liao & Huebner 2020, p. 9). We might refer to 'whites only' signs (as well as the Shirley Cards discussed in Chapter 2) as 'racist scaffolding': external structures that change the cognitive demands of tasks in ways that are congruent with a racially oppressive system—undermining one group while benefiting another. This could be a subclass of hostile scaffolding and additional research could consider how interests are undermined and served, how beneficiaries are causally 'in the loop', as well as whether some instances of 'racist scaffolding' are deep.

### *Gamification in the Workplace (Uber, Disney & Amazon)*

In §3.4, I posited that Twitter gamifies communication with deep hostile scaffolding (Nguyen, 2021). Other companies have employed similar techniques to increase worker productivity. Uber, for example, gamifies driving by offering drivers scores and achievements (Nguyen, 2020). Much like what we saw in Slaby (2016) and Nguyen (2021), a driver's working values may be negatively shifted—resulting in longer working hours. One way to achieve this is through nudges[64] (techniques that increase users' time on device). When Uber drivers wish to go offline, a gamified nudge shows a needle with an arbitrary goal—attempting to incentivise the driver to work longer (Narayanan et al., 2020). Other harmful gamification systems have also been implemented by

[64] This is also an example of a dark pattern, which will be discussed in the section below.

Disney and Amazon. Disney implemented a system that tracked the performance of laundry staff by posting real-time leaderboards, which displayed each staff member's number of washed and folded sheets, towels and comforters (Nguyen, 2020). Names were then presented in green, yellow or red—depending on individual performance and productivity. As a result, staff members began to compete with one another, driving increased productivity rates. However, staff members also began engaging in behaviour that was more detrimental to their welfare—such as missing bathroom breaks or taking actions that were more likely to result in injury (Nguyen, 2020). They noted how they found it difficult to ignore "the motivational pull of the game-like elements"—referring to the system as "the electronic whip" (Nguyen, 2020, p. 200). Similarly, Amazon has also used scoring techniques to motivate warehouse staff, which have led to comparable safety concerns—since employees who lagged behind were also more likely to take risks (Anderson, 2021).

These scoring systems act as motivational scaffolding while also displaying deep qualities—since the external structures perform real-time assessments of individual performance. Information is then broadcast back to workers, which shapes behaviour in ways that undermine their interests—overriding previously healthy work goals. The resulting productivity then benefits other agents. There is room for the relationship between hostile scaffolding and gamification to be expanded on in future projects.

### *Virtual Reality, Dark Patterns & AI*

As demonstrated above, developing technologies provide us with new challenges and a hostile scaffolding framework may offer ways to better understand future issues. Virtual reality (VR) is another rapidly developing technology. Headsets can display immersive virtual spaces and track the user's real-world movements and positions in space. Virtual classrooms, for example, may use techniques that ensure that a teacher's avatar constantly maintains eye contact with each student in the class—resulting in improved attention (Bailey, 2016). In §3.2, I discussed sunglasses as shallow hostile scaffolding when they obfuscated facial cues (Viola, 2022). VR technology could take this a step further. When engaged in social VR interactions, software may alter facial features, an agent's appearance as well as their gestures in order to influence others in harmful ways (Bailey, 2016). This could take the form of software mimicking or manipulating body language and facial expressions so as to create more favourable impressions. When used by agents to deceive and profit (suppose there was a deceptive virtual salesperson), this could become hostile scaffolding (perhaps even deep). Similarly, Tseng et al. (2022) explore techniques that they refer to as *Virtual-Physical Perceptual Manipulations* (VPPMs). These are "Extended Reality

(XR) driven exploits that alter the human multi-sensory perception of our physical actions and reactions to nudge the user's physical movements", and are often imperceivable (Tseng et al., 2022, para. 1). For example, a user's movement path may be slightly rotated in order to steer them in certain directions (Tseng et al., 2022). As noted in §3.3.1, the more control one agent has over an environment, the more they can manipulate it in ways that can benefit them. Casinos already alter spaces to direct patrons' walking patterns (Friedman, 2000) and VR environments could use similar techniques to orient users in virtual spaces. In settings where every element of an environment is controlled (and used by multiple agents), hostility is likely. Tseng et al. (2022) posit that it may be possible for hackers to install malware onto VR headsets which could result in harm—such as software that deliberately manipulates agents into walking into real world objects. But other potential manipulations that benefit agents (such as advertisers) might also deserve attention. For example, imagine walking through a virtual store where all the products have been chosen and positioned in ways that increase the likelihood that you will purchase them.[65] Additional research into hostile scaffolding and VR could analyse these types of situations.

Other areas where hostile scaffolding could be considered include the use of dark patterns. These are user interfaces that manipulate agents into performing actions that they did not intend to do—such as purchasing or subscribing to something (Mathur et al., 2021). Brignull (n.d.) introduces 12 types of dark patterns. These include, to name a few: Bait-and-Switch techniques (where "the user sets out to do one thing, but a different, undesirable thing happens instead"); Confirmshaming ("the act of guilting the user into opting in to something"); Misdirection (where "the design purposefully focuses your attention on one thing in order to distract your attention from another"); and Privacy Zuckering (where "you are tricked into publicly sharing more information about yourself than you really intended to") (Brignull, n.d.). Some of the language used to describe these techniques is similar to the characterisation of hostility laid out in Chapter 2, with some describing dark patterns "as subverting user intent" or "or subverting user preferences" (Mathur et al., 2021). Furthermore, they are often implemented with the intention of achieving the goals set by designers—thus placing an agent causally 'in the loop' (Mathur et al., 2021; Zagal et al., 2013). Similarities can also be drawn with the gambling techniques described in Chapter 3, with some computer and mobile games implementing dark patterns that lead to players losing track of how much they spend (Zagal et al., 2013; Lewis, 2014). When considering dark patterns as a whole, the harming of an agent's welfare appears to

[65] This, of course, already happens in real-world stores, but VR technologies could develop more individualised, adaptive and manipulative techniques.

be common. Mathur et al. (2021) discuss potential harms such as *financial loss*, the *invasion of privacy*, and increasing the cognitive burden of tasks. Additionally, they argue that "the vast majority of dark patterns attempt to undermine individual autonomy" in some way (Mathur et al., 2021, p. 19). For these reasons, dark patterns are, at least *prima facie*, congruent with shallow hostile scaffolding (and some cases could even be deep). Future hostile scaffolding work could aid in developing the discourse around dark patterns.

Finally, a hostile scaffolding lens could also be applied to artificial intelligence (AI). Hernández-Orallo and Vold's (2022) work greatly overlaps with the complementarity thesis (as discussed in §1.5). They argue that certain external tools that use AI are extended cognitive processes when "sufficiently tightly coupled with a person's cognitive system" (Hernández-Orallo & Vold, 2022). For Hernández-Orallo and Vold (2022), these are not only "cognitive assistants" or "cognitive prosthetics" (although there is overlap), but can be parts of an agent's mind (p. 11). Examples include devices using AI programs such as COACH (Cognitive Orthosis for Assisting Activities in the Home), that monitor and assist people with dementia, or virtual cognitive behaviour therapists, such as Tess and Ellie (Hernández-Orallo & Vold, 2022, p. 12). However, unlike other kinds of cognitive extenders (like Otto's notebook), external tools using machine learning (and other AI functionalities) are "qualitatively different" from more typical extenders (Hernández-Orallo & Vold, 2022, p. 9). Hernández-Orallo & Vold (2022) describe how "these systems can perceive, navigate, make complex decisions, recognize and produce language, plan, identify emotions, etc., all in complex and changing situations" (p. 9). I describe deep scaffolding as facilitating (and transforming) significant cognitive work and these types of AI extenders appear to fit this description.[66] Not only can more work be done to better analyse how AI extenders work as deep benign scaffolds (which assess and treat cognitive disorders) but we should also account for how future AI developments may exploit individuals with cognitive impairments or vulnerabilities. Hernández-Orallo & Vold (2022) list five risks: *autonomy*, *reliability*, *unregulated or recreational use*, *moral status* and *privacy and allowance in the public* (p. 31). I will only focus on the first risk here, *autonomy*—since it is most relevant to hostile scaffolding. When AI extenders use nudges and interventions (based on user information), they could "effectively bypass the agent's right to decide" and encourage certain actions "without appealing to the agent's rationality (e.g. by presenting them with reasons to act)" (Hernández-Orallo & Vold, 2022, p. 31). Hernández-Orallo & Vold (2022) posit that "these devices could risk becoming

[66] These also display a level of adaptability that other deep cases (such as a calendar) do not.

*manipulative*" (p. 31). For example, pivoting briefly to a helpful case, possible benign AI scaffolds include machine learning that could aid individuals suffering from addiction:

> So for a particular person, a machine learning system could detect that the person is likely to have depleted self-control when meeting with certain friends or going to certain places where they used to smoke. Suggesting walking routes that avoid smoking zones, or even reminding the agent that these situations may be challenging (and perhaps directing them to resources) might help them in controlling their impulses. (Hernández-Orallo & Vold, 2022, pp. 19-20)

A scenario could be imagined where a machine learning system (perhaps by design or sneakily installed malware) exploited users by recommending walking routes that increased their likelihood to purchase certain products or engage in certain activities. This could (potentially) be equivalent to a deep implementation of the shallow casino interior scaffolds described in §3.3.1. I argue that hostile scaffolding offers useful criteria to start identifying these cases.

## 4.2. Conclusion

Let us conclude by reviewing the main moves and arguments made throughout this dissertation. In Chapter 1, I discussed and expanded on the scaffolding literature (Hutchins, 1995a; Clark, 1997; Sterelny 2003, 2010). I then drew a distinction between deep and shallow scaffolding and argued that, under the complementarity thesis (Menary, 2006; Sutton, 2010), cases of deep scaffolding were extended cognitive processes. I concluded the chapter by highlighting how scaffolding is often viewed as 'benign' (or 'helpful') and suggested that hostile scaffolding is also possible. In Chapter 2, I drew from Sterelny (2003) to expand on 'hostility'—characterising it as a property of an environment, relative to an agent, that undermines the interests of certain agents, while serving those of others. I then discussed 'interests' in more detail and showed how hostile scaffolding could be separated from harmful scaffolding by asking "Who suffers?" and "Who benefits?", as well as placing a beneficiary casually 'in the loop'. I then re-emphasised how much of our activity was defined by scaffolding (Clark, 2003; Hutchins, 1995a, 1995b; Kirsh, 1996; Sterelny, 2010); and discussed how our potential vulnerabilities are often not properly addressed. I concluded Chapter 2 by reviewing other thinkers who shared similar concerns to my own (Aagaard, 2020; Liao & Huebner, 2020; Slaby, 2016). However, it was shown that the notion of hostile scaffolding (and deep scaffolding) was not considered in the same way as this project articulates. Chapter 3's aim was then to provide genuine hostile scaffolding examples. I provided shallow examples in the form of sunglasses (Viola, 2022), casino floor design and casino ambience (Friedman, 2000; Schüll, 2012). Most crucially, I provided real world cases of deep hostile scaffolding through my analysis of gambling machines that employ tracking-based interventions and virtual reel mapping (Schüll, 2012), as well as Twitter's gamification techniques (Nguyen, 2021). These cases demonstrate that deep hostile scaffolding is a real phenomenon and should be taken seriously. My criteria also successfully identified each case as shallow or deep hostile scaffolding. Finally, since hostile scaffolding is still a new project, §4.1 briefly suggested other areas where this work could be expanded. Overall, this dissertation expanded on the scaffolding literature in useful ways (by considering hostility), showed that hostile scaffolding is real and relevant, and highlighted other areas where this work could be developed.

# Reference List


Aagaard, J. (2020). 4E cognition and the dogma of harmony. *Philosophical Psychology*, *34* (2), 165-181. https://doi.org/10.1080/09515089.2020.1845640

Abbott, M. (2006). Do EGMs and Problem Gambling Go Together like a Horse and Carriage? *Gambling Research*, *18*, 7-38.

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*, 43-64. https://doi.org/10.1080/09515080120033571

Adams, F., & Aizawa, K. (2006). Defending the bounds of cognition. In R. Menary (Ed.), T*he Extended Mind* (pp. 45-48). Ashgate.

Anderson, G. (2021, May 23). Amazon Gamifies Warehouse Worker Tasks, Raising Experts' Concerns. *Forbes*. https://www.forbes.com/sites/retailwire/2021/03/23/amazon-gamifies-warehouse-worker-tasks-raising-experts-concerns/?sh=c1c9bd03810f

Bailey, D. E. (2016, August 25). How virtual reality can manipulate our minds. *Phys.org*. https://phys.org/news/2016-08-virtual-reality-minds.html

Beer, R.., & Chiel, H. (1993). Simulations of cockroach locomotion and escape. In R. Beer, R. Ritzmann, T. McKenna (Eds.), *Biological Neural Networks in Invertebrate Neuroethology and Robotics* (pp. 267-286). Academic Press.

Bernstein, N. A. (1947). *O postroenii dvizheniy* [On the construction of movements]. Medgiz.

Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. Pergamon Press.Bickle, J. (2020). Multiple Realizability. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 ed.). Stanford University. https://plato.stanford.edu/archives/sum2020/entries/multiple-realizability

Blaszczynski, A. (2008). *Expert Report of Professor Alex Blaszczynski: In the Matter of Jean Brochu v. Loto Québec et al. — Class action.* media.cleveland.com/metro/other/Blaszczynski%20expert%20deposition%20on%20slots%20addictiveness.pdf

Buckner, C., & Garson, J. (2019). Connectionism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition). Stanford University. https://plato.stanford.edu/archives/fall2019/entries/connectionism/

Braddon-Mitchell, D., & Jackson, F. (2007). *The Philosophy of Mind and Cognition* (2nd ed.). Basil Blackwell.

Brignull, H. (n.d.). Types of deceptive design. *Deceptive Design*. https://www.deceptive.design/types

Brooks, R. A. (1990). Elephants Don't Play Chess. *Robotics and Autonomous Systems*, *6*, 3-15. https://doi.org/10.1016/S0921-8890(05)80025-9

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*(1-3), 139-159. doi:10.1016/0004-3702(91)90053-m

Carter, J. A., & Palermos, S. O. (2016). Is having your computer compromised a personal assault? The ethics of extended cognition. *Journal of the American Philosophical Association*, *2*(4), 542-560. doi:10.1017/apa.2016.28

Clark, A. (1997). *Being There: Putting Mind, Brain and Body Together Again*. MIT Press.

Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence. Oxford University Press.*

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7-19. doi:10.1093/analys/58.1.7

Clowes, R. (2019). Immaterial engagement: Human agency and the cognitive ecology of the internet. *Phenomenology and the Cognitive Sciences*, *18*(1), 259-279. https://doi.org/10.1007/s11097-018-9560-4

Cole, M., Hood, L., & McDermott, R. (1978). Ecological niche picking: Ecological invalidity as an axiom of experimental cognitive psychology. In U. Neisser (Ed.), *Memory Observed: Remembering in Natural Contexts*. W. H. Freeman. doi:10.13140/2.1.4727.1204

Collins. (n.b.). Scaffolding. In *Collins Online Dictionary*. Retrieved August 8, 2021, from https://www.collinsdictionary.com/dictionary/english/scaffolding

Colombetti, G., & Kruger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology*, *28*(8), 1157-1176. doi:10.1080/09515089.2014.976334

Del Barco, M. (2014, November 18). How Kodak's Shirley cards set photography's skin-tone standard. *NPR*. https://www.npr.org/2014/11/13/363517842/for-decades-kodak-s-shirley-cards-set-photography-s-skin-tone-standard

Dennett, D.C. (2006). *Breaking the spell: Religion as a natural phenomenon.* Viking.

Dotson, T. (2012). Technology, choice and the good life: Questioning technological liberalism. *AI & Society*, *34*(4), 326-336. doi: 10.1016/j.techsoc.2012.10.004

Donald, M. (1991). *Origins of the Modern Mind*. Harvard University Press.

Fields, K. E., & Fields, B. J. (2014). *Racecraft: The soul of inequality in American life*. Verso.

Friedman, B. (2000). *Designing Casinos to Dominate the Competition*. Institute for the Study of Gambling and Commercial Gaming.

Giacalone, L. D., Jr. (1996). *Dynamic rate control method and apparatus for electronically played games and gaming machines*. (U.S. Patent No. 5758875). U.S. Patent and Trademark Office. https://patents.google.com/patent/US5758875A/en

Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford University Press.

Gregg, M. (2011). *Work's Intimacy*. Polity.

Griffiths, M. D. (2009). Casino Design: Understanding Gaming Floor Influences on Player Behaviour. *Casino and Gaming International, 5*(1), 21-26.

Griffiths, P., & Scarantino, A. (2009). Emotions in the wild: The situated perspective on emotion. In M. Aydede, & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 437-453). Cambridge University Press.

Harrigan, K. A. (2007). Slot Machine Structural Characteristics: Distorted Player Views of Payback Percentages. *Journal of Gambling Issues*, 20, 215-234.

Hirsh, A. R. (1995). Effects of Ambient Odors on Slot- Machine Usage in a Las Vegas Casino. *Psychology & Marketing, 12*(7), 585-594. https://doi.org/10.1002/mar.4220120703

Hernández-Orallo, J., & Vold, K. (2019). AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. *AIES'19: AAAI / ACM Conference on AI, Ethics, and Society*, Honolulu, HI, USA. https://doi.org/10.1145/3306618.3314238

Hollan, J.D., Hutchins, E.L., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput. Hum. Interact., 7(2)*, 174-196. doi: 10.1145/353485.353487

Hölldobler, B. & Wilson. E. (1990). *The Ants*. Harvard University Press.

Holtmann, A. (2004). The Sound of Music: Hi- Tech Audio Systems Are Giving Casinos a Wider Variety of Musical Offerings to Choose From; and More Control over Them. *Casino Journal*, July, 3-49.

Hurley, S. L. (2006). Varieties of externalism. In R. Menary (Ed.), *The Extended Mind (pp. 101-153.).* MIT Press. doi:10.7551/mitpress/9780262014038.003.0006

Hutchins, E. (1995a). *Cognition in the Wild*. MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speed. *Cogn. Sci.*, *19*, 265–288. https://doi.org/10.1016/0364-0213(95)90020-9

Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics*, *37*, 1555-1577. https://doi.org/10.1016/j.pragma.2004.06.008

Kaufmann, J. C. (2011). *Le sac: Un petit monde d'amour*. J. C. Latte`s.

Kirsh D. (1996). Adapting the environment instead of oneself. *Adapt. Behav.*, *4* (3-4), 415-452. https://doi.org/10.1177/105971239600400307

Kirsh, D. & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, *18*, 513-549. https://doi.org/10.1016/0364-0213(94)90007-8

Kiverstein, J. (2018). Extended Cognition. In A. Newen, L. Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E-cognition* (pp. 19-40). Oxford University Press.

Kim, G., Seong, S. H., Hong, S. S., & Choi, E. (2022). Impact of face masks and sunglasses on emotion recognition in South Koreans. *PloS One*, *17*(2), e0263466. https://doi.org/10. 1371/journal.pone.0263466

Kuhn, R. L [Closer To Truth]. (2020). What is Extended Mind? | Episode 1811 | Closer To Truth [Video]. YouTube. https://www.youtube.com/watch?v=rSngHyD44OY

Levin, J. (2018, *July 20).* Functionalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Stanford University. https://plato.stanford.edu/archives/win2021/entries/functionalism

Lewis, C. (2014). Monetary Dark Patterns. In C. Lewis (Ed.), *Irresistible Apps: Motivational Design Patterns for Apps, Games, and Web-based Communities* (p.111-117). Apress. https://doi.org/10.1007/978-1-4302-6422-4 10

Liao, S. & Huebner, B. (2020). Oppressive Things. *Philosophy & Phenomenological Research, 103*(1), 92-113. doi:10.1111/phpr.12701

Lloyd, J. E. (1997). The Evolution of Mating Systems in Insects and Arachnids: Firefly mating ecology, selection and evolution. In J. C. Choe, & B. J. Crespi (Eds.), *Mating Systems in Insects and Arachnids* (pp. 184-192). Cambridge University Press. doi:10.1017/CBO9780511721946.011

Luria, A. R., & Vygotsky, L. S. (1992). The child and his behavior. In *Ape, primitive man, and child: Essays in the history of behaviour* (E. Rossiter, trans.). Harvester Wheatsheaf. https://www.marxists.org/archive/vygotsky/works/1930/man/index.htm

Mathur, A., Kshirsagar, M., & Mayer, J. (2021, May 8-13) *What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods*. CHI '21: CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. https://doi.org/10.1145/3411764.3445610

McLeod, S. A. (2019). What Is the zone of proximal development?. *Simply Psychology*. www.simplypsychology.org/Zone-of-Proximal-Development.html

Menary, R. (2006). Attacking the Bounds of Cognition. *Philosophical Psychology, 19*, 329-344. doi:10.1080/09515080600690557

Menary, R. (2018). Cognitive Integration How Culture Transforms Us and Extends Our Cognitive Capabilities. In A. Newen, L. Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognitio*n (pp. 187-215.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.10

Narayanan, A., Mathur, A., Chetty, M., & Kshirsagar, M. (2020). Dark Patterns: Past, Present, and Future: The evolution of tricky user interfaces. *Queue*, *18* (2), 67-92. https://doi.org/10.1145/3400899.3400901

Newen, A., de Bruin, L., & Gallagher, S. (2018). 4E Cognition. Historical Roots, Key Concepts, and Central Issues. A. Newen A., L. Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E-cognition* (pp. 3-15). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780198735410.013.1

Nguyen, C. T. (2020). *Games: Agency as Art*. Oxford: Oxford University Press.

Nguyen, C. T. (2021). How Twitter gamifies communication. In J. Lackey (ed.), *Applied Epistemology* (pp. 410-436*)*. Oxford University Press. doi:10.1093/oso/9780198833659.003.0017

Piccinini, G., & Scarantino, A. (2010). Computation vs. information processing: Why their difference mattersto cognitive science. *Studies in History and Philosophy of Biological and Biomedical Science*, *41*, 237-246. https://doi.org/10.1016/j.shpsa.2010.07.012

Piredda, G. (2019). What is an affective artefact? A further development in situated affectivity. *Phenomenology and the Cognitive* Sciences, *19*(2), 1-19. https://doi.org/10.1007/s11097-019-09628-3

Roth, W.-M., & Jornet, A. (2017). *Understanding educational psychology: A late Vygotskian, Spinozist approach*. Springer.

Ross, D. (2020). Addiction is socially engineered exploitation of natural biological vulnerability. *Behavioural Brain Research*, *386*, 112598. https://doi.org/10.1016/j.bbr.2020.112598

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, *101*, 389-428. doi:10.5840/jphil2004101826

Schüll, N.D. (2012). *Addiction by Design: Machine Gambling in Las Vegas*. Princeton University Press.

Skinner, B. F. (1953). *Science and human behavior*. Macmillan.

Slaby, J. (2016). Mind Invasion: Situated Affectivity and the Corporate Life Hack. *Frontiers in Psychology*, *7*:266. doi:10.3389/fpsyg.2016.00266

Shapiro, L., & Spaulding, S. (2021, June 25). Embodied Cognition. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Stanford University. https://plato.stanford.edu/archives/win2021/entries/embodied-cognition/

Shvarts, A., & Bakker, A. (2019). The early history of the scaffolding metaphor: Bernstein, Luria, Vygotsky, and before. *Mind, Culture, and Activity*, *26*(1), 4-23. doi:10.1080/10749039.2019.1574306

Stephan A. (2018). 3 E's are sufficient, but don't forget the D. In A. Newen, L. Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E-cognition (*pp. 607-619). Oxford University Press. https://doi.org/10.1177/1754073922111241

Sterelny, K. (2003). *Thought in a Hostile World*. Oxford: Blackwell.

Sterelny, K. (2010). Minds: extended or scaffolded? *Phenomenology and the Cognitive Sciences*, *9*(4), 465-481. doi:10.1007/s11097-010-9174-y

Sutton, J. (2010). Exograms, interdisciplinarity and the cognitive life of things. In R. Menary (Ed.), *The Extended Mind* (pp. 189-225). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0009

Thompson, I. (2009, January 7). Meet Your New Neighbor: How Slot Machines Are Secretly Designed to Seduce and Destroy You, and How the Government Is in on It. *Philadelphia City Paper*. http://stoppredatorygambling.org/wp-content/uploads/2012/12/Meet-Your-New-Neighbor.pdf

Telnaes, I.S. (1984). *Electronic gaming device utilizing a random number generator for selecting the reel stop positions.* (U.S. Patent No. 4448419*)*. U.S. Patent and Trademark Office. https://patents.google.com/patent/US4448419A/en

Tribble, E. (2005). Distributing cognition in the globe. *Shakespeare Quarterly*, *56*, 135-155. doi: 10.1353/shq.2005.0065

Tseng, W.-J., Bonnail, E., McGill, M., Khamis, M., Lecolinet, E., Huron, S., & Gugenheimer, J. (2022, April 29-May 5). *The dark side of perceptual manipulations in virtual reality*. CHI '22: CHI Conference on Human Factors in Computing Systems. New Orleans, LA, USA. https://doi.org/10.1145/3491102.3517728

Valsiner, A. (1987). *Culture and the Development of Children's Action*. John Wiley & Sons Inc.

Viola, M. (2022). Seeing through the shades of situated affectivity. Sunglasses as a socio-affective artifact. *Philosophical Psychology,* forthcoming. https://doi.org/10.1080/09515089.2022.2118574

Walter, S. (2018). Critical Note: Cognition, Action, and Self-Control from the 4E Perspective. A. Newen, L. Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E-cognition* (pp. 321-332). Oxford University Press.

Webb, B. (1994). Robotic experiments in cricket phonotaxis. In D. Cliff, P. Husbands, J. A. Meyer, & S. Wilson (Eds.), From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior (pp. 45–54). MIT Press.

Wheeler, M. (2010). In Defense of Extended Functionalism. In R. Menary (Ed.), *The Extended Mind* (pp. 245-270). MIT Press. doi:10.7551/mitpress/9780262014038.003.0011

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89-100. doi:10.1111/j.1469-7610.1976.tb00381.x

Zagal, J. P., Björk, S., & Lewis, C. (2013, May 14-17). *Dark Patterns in the Design of Games*. FDG 2013: Foundation of Digital Games 2013, Chania, Crete, Greece. https://core.ac.uk/download/pdf/301007767.pdf