# Technology and the Space of Reasons:
# A Pragmatist Philosophy of Synthetic Media

by

Asher Brandon Caplan

A dissertation accepted and approved in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Philosophy

Dissertation Committee:
Colin Koopman, Chair
Mark Johnson, Core Member
Ramón Alvarado, Core Member
Bryce Newell, Institutional Representative

University of Oregon

Spring 2026

© 2026 Asher Brandon Caplan

This work is openly licensed via CC BY 4.0.



# DISSERTATION ABSTRACT

Asher Brandon Caplan

Doctor of Philosophy in Philosophy

Title: Technology and the Space of Reasons: A Pragmatist Philosophy of Synthetic Media

Technological artifacts play an important but under-appreciated role within practices of justification. In light of the emerging epistemic challenges posed by AI-generated content from deepfake disinformation to agentic chatbots, it is crucial that we better understand how artifacts feature in the justificatory practice of what pragmatist philosophers have called the space of reasons.

The first part of "Technology and the Space of Reasons" offers an internal critique of the linguistic account of normativity and justification found in the work of pragmatist philosophers Wilfrid Sellars, Richard Rorty, and Robert Brandom. According to these philosophers, justification is modeled according to a game of giving and asking for reasons where linguistic interlocutors trade spoken or written sentences with each other. In the standard account, this practice of justification is a conversational affair, and performances which are not explicitly articulated in a sentential form do not count as relevant moves within the game. Against this linguistic account of the space of reasons, I draw from recent work in philosophical pragmatism to defend a more expansive space of reasons, one in which participants also deploy artifactual reasons. In these cases, non-linguistic artifacts such as photographs, thermometers, and the various features of online platforms are socially indexed to adjudicate epistemic disagreements.

The second part of the project applies the theory developed in part one to two problems posed by AI-generated synthetic media: the player question and the object question. The player question concerns who or what participates in the game of giving and asking for reasons. To answer this question, I argue that current conversational AI agents should not be considered fellow reasoners, since the space of reasons includes practices of justification that utilize abilities well beyond mere conversational engagement. The object question concerns what entities count as reasons. In relation to this question, I apply the account of artifactual reasons to track our current practices of verification for audio-visual sources in light of the disruptions posed by deepfakes. Once deepfake verification is understood as a practice of social justification involving multiple

competing audio-visual sources, concerns about the potentially catastrophic epistemological effects of deepfakes are substantially mitigated.

# ACKNOWLEDGMENTS

This dissertation would not have been possible without several fellow players in the game of giving and asking for reasons. First and foremost, I would like to thank my committee for their interest and attentiveness toward my project. Mark Johnson taught me the very best of the classical pragmatists and graciously agreed to serve as a committee member in retirement. Ramón Alvarado encouraged me from the onset to distinguish myself by diving into the nitty-gritty technical details of large language models to the betterment of the project. Bryce Newell generously read and commented on early drafts of the fifth chapter and welcomed me into his professional network on surveillance studies. Finally, Colin Koopman has offered everything that I could ask for from an advisor. I thank him for letting me chart out my own path within these debates that he, himself, knows so well. I truly think this dissertation is a product of the pluralism and effective interdisciplinary engagement that he encourages and enacts so well.

Individual chapters within the dissertation were informed by extensive discussions with colleagues and mentors, both formal and informal, held during the past five years. Part I benefited from extensive conversations with my fellow Rortyans at meetings of the Richard Rorty Society and the Society for the Advancement of American Philosophy. I would like to especially thank David Rondel, Paul Showler, Elizabeth Cantalamessa, and Chris Voparil for a series of lively debates. Part II, especially my account of deepfakes, is informed by discussions with Tobias Flattery and Megan Hyska and from a generous read-through by Heather Douglas at the *Science, etc. Graduate Conference* at the University of Washington. Gabriela Martinez and Hannah Waight also reviewed early drafts of the deepfakes chapter and provided insightful feedback from beyond the discipline of philosophy. I also extend gratitude to Kevin Driscoll and Christopher Ali for advising me in my undergraduate and early graduate career in media studies. This project certainly took shape under their guidance and still bears the mark of the Department of Media Studies at the University of Virginia.

Lastly, I would like to acknowledge the support provided by my friends and family over the course of my graduate career. I thank several fellow PhD students for reading and engaging with prior drafts of the chapters, including Abbas Bagwala, Gonzalo Bustamante-Moya, Brooke Burns, Matt Miyagi Tuten, Rugile Kasperiunaite, Chelsea Schwartz, and MacKenzie Brinn. In addition to this list, Professors Nicolae Morar and Jacob P. Neal fostered a unique space to think through pragmatism and its relation to the philosophy of science and have cheered me on throughout the

process. Moaaz Alqady and Hunter Clark both brought helpful perspectives outside of philosophical pragmatism during discussions about the opening chapters. Others assisted the dissertation writing process primarily through their comradery and friendship. In Eugene, this group includes, but is not limited to, Amanda Dubrule, Maia Wellborn, Joseph Dillabough, Augie Schiano, Audrey Brown, Edgar Llamas, Matt Stone, Kirstin Waldkoenig, Sophia Spangler, Randall Rodrigez, and Avi Locke. Back home in Virginia, it includes Sam and Stephanie Roche, Jeff Niznik, Kaitlyn Elliott, Aaron and Kim Caplan, and last but not least, my parents, who have loved and supported me in spite of my unusual decision to become a philosopher.

For Atlas, Alina, and Adaline.

# TABLE OF CONTENTS


Chapter 1 – Introduction ..... 11

1.1 – Technology as Epistemology ..... 11
1.2 – The Problems of Synthetic Media ..... 11
1.3 – Reasons, Norms, and The Manifest Image ..... 13
1.4 – Practicing Pragmatism ..... 17
1.5 – Reasoning with Synthetic Media ..... 21

PART I – Reasons and Artifacts in Contemporary Pragmatism ..... 25

Chapter 2 – Artifactual Pragmatism: On Reason and Practice After Rorty ..... 26

2.1 – Introduction ..... 26
2.2 – Rorty's Linguistic Turn ..... 28
2.3 – Pragmatism After Rorty: Homo Sapiens or Homo Faber? ..... 36
2.4 – From Neo-pragmatism to Artifactual Pragmatism ..... 46
2.5 – An Outline of Artifactual Pragmatism ..... 50
2.6 – A Puzzle about Reasons: Signs, Artifacts, and the Matching Game ..... 58

Chapter 3 – Reasons Reconsidered ..... 63

3.1 – Introduction ..... 63
3.2 – Slabs, Slaps, and Photographs: The Scope of Conversation ..... 65
3.3 – Rehabilitating Evidence: Triangulation and Perceptual Content ..... 68
3.4 – Artifactual Reasons ..... 85
3.5 – Conclusion ..... 93

PART II – Pragmatism and Synthetic Media ..... 95

Chapter 4 – Alien Reasoners: Triangulation, Multimodality, and Conversational AI ..... 96

4.1 – Introduction ..... 96
4.2 – Player or Object? ..... 99
4.3 – Triangulation and Analytic Pragmatism ..... 106
4.4 – Triangulating Reasons ..... 112
4.5 – Conclusion: Pragmatism and the Normative Status of LLMs ..... 117

Chapter 5 – Deepfake Detection and Artifactual Deliberation ..... 120

5.1 – Introduction ..... 120

5.2 – The Epistemological and Ethical Harms of Deepfakes .....121
5.3 – Deepfake Detection and the Epistemic Arms Race .....126
5.4 – Artifactual Deliberation .....129
5.5 – Conclusion: The Deep and Shallow Harms of Deepfakes .....142

Chapter 6 – Conclusion: Technology, Normativity, and the Pragmatist Enlightenment .....145

Bibliography .....152

# FIGURES

Figure 1: An example of how LLMs generate text through next-word prediction ..... 101
Figure 2: An LLM supplemented with multimodal capabilities ..... 106
Figure 3: A comparison of a deepfake video with a corresponding authentic video ..... 140

# CHAPTER 1:
# INTRODUCTION

> As a genius of construction man raises himself far above the bee in the following way: whereas the bee builds with wax that he gathers from nature, man builds with the far more delicate conceptual material which he first has to manufacture from himself.
>
> — Friedrich Nietzsche, "On Truth and Lies in a Nonmoral Sense" (1873)

## 1.1 – Technology as Epistemology

The history of ideas is bound up with the history of technology. From Galileo's telescope to Boyle's air pump and from atom bombs to perceptrons, many of the most far-reaching advancements in human knowledge have been attained as part of a complex combination of intellectual matter with actual matter, a mixture which is most directly encountered in our use of artifacts. These objects are not only related to each other in a causal manner, an ebb and flow of cascading effects through space and time. They also interact within a rational fabric. When artifacts clump up, combine, and disperse into differing arrangements, they open up new epistemic environments, each providing differing resources for reasoning. And so, the history of technology may also be understood as a history of epistemology, one where new artifactual additions bring new modes, practices, and objects into our practices of giving reasons.

## 1.2 – The Problem of Synthetic Media

Generative AI is the latest technology to radically reweave the rational fabric. In 2020, philosophers Regina Rini (2020) and Don Fallis (2020) separately sounded the alarm concerning a new class of hyper-realistic audio-visual deepfakes. These forms of AI-generated "synthetic media" have the potential to undermine trust in all forms of indirect evidence and thus, they threaten to uproot from our epistemic landscape those recording technologies which have protected our testimonial norms from erosion.

In connection to these fears of informational collapse, new kinds of excusatory and exculpatory practices are also enabled by this class of synthetic media. Legal scholars Robert

Chesney and Danielle Citron (2019) point to instances where unscrupulous politicians may challenge compromising audio or video recordings by suggesting that they are AI-generated, what they call the "liar's dividend." And in recent years, these socio-political concerns have broadened to consider how the widespread adoption of generative AI will modify both internal state propaganda (Schoenherr 2023, Rose 2026) and international disinformation campaigns (Allyn 2022; Jones 2026).

Text-based synthetic content, such as that generated by ChatGPT, also brings forth a unique set of concerns. Given that the proliferation of synthetic text-based content destabilizes trust in the printed word, this content should not be excluded from analysis. However, since much of this synthetic text is produced through human collaboration with conversational AI agents, synthetic textual content is often framed differently from audio-visual media. It is treated in terms of the underlying generative model rather than in terms of the produced content. With this shifted emphasis in play, synthetic media may be investigated as an even more unusual kind of addition to the artifactual environment, something which is quasi-agential in character.

Many scholars have contested this sort of treatment of text-based AI. Yet, while some have labeled conversational AI systems "stochastic parrots" which merely repeat inherently meaningless statistical patterns present within their training data (Bender et al. 2021) and others have taken up even stronger language in treating them as peddlers of "bullshit" (Hicks et al. 2024: Fisher 2024), there is no denying that current conversational AI produce text which resembles authentic meaning-making to an impressive degree. A range of empirical studies show that synthetic text-based content can fool participants of various backgrounds into believing the content is authentic human writing (Gao et al., 2023; Jones and Bergen 2024; Waltzer et al. 2024).[1] Even more impressive, two thirds of a surveyed population were willing to grant some phenomenal consciousness to an artificial conversational agent (Colombatto and Flemming 2024), suggesting that, though critics deny intentionality to conversational AI, many lay persons see the increasing complexity of textual responses as a proxy for attributing sentience to artificial systems. These attributions, whether they are erroneous or not, contribute to a broader philosophical discussion on the moral status of artificial systems (see Chalmers 2023), one which no longer depends solely upon science fiction scenarios.

[1] This is not to mention instances of "false positives" where authentic text is incorrectly identified as synthetic.

These two variations on the problem of synthetic media, namely, deepfakes and conversational AI, form the practical background of this dissertation. Debates concerning synthetic textual content track issues on how conversational AI, generally presumed to be large language models (LLMs), undermine the standard means by which the *participants* of a communicative activity have been demarcated. These debates are connected to a class of philosophical questions regarding the agency, responsibility and normative status of LLM-enabled conversational AI systems. Epistemological and political analyses of deepfake disinformation track a different set of debates concerning *evidence* and its associated quandaries. They are connected to a class of philosophical questions regarding the epistemic status of non-sentential content that is neither structured like a belief (a sentential attitude) nor a fact (a state of affairs).

### 1.3 – Reasons, Norms, and the Manifest Image

One powerful philosophical metaphor which connects the two sites of practical investigation to the abstract debates on normativity and epistemic content is the *space of reasons,* a term which describes the social practice of giving and asking for reasons among linguistically-endowed interlocutors. First introduced in the seminal essay "Empiricism and the Philosophy of Mind" by Wilfrid Sellars ([1956] 1997), this metaphor initially served as part of an extended attack on the idea that knowledge has empirical foundations. In Sellars' original formulation, the space describes how knowledge requires that perceptions be supplemented by a broader social and rational edifice. They are to be placed in "the logical space of reasons, of justifying and being able to justify what one says" (Sellars [1956] 1997, 76).

Although Sellars' original paper is more concerned with challenging epistemic appeals to some immediately "given" empirical content, Robert Brandom (1997), helpfully shifts the emphasis of the metaphor to that of persons, agents, or performers rather than their perceptual episodes. In Brandom's reconstruction, some entity is said to be in the space of reasons when it engages in a particular social practice or language-game, *the game of giving and asking for reasons*. Accordingly, the organization of the mental is variously entangled with the organization of the social; humans can only be rational animals when they are also social animals. Without an interlocutor to give reasons *to* or ask reasons *from*, there are no reasons whatsoever. And so, the metaphor of the "game" singles out the mutual dependency of the players. Each player's very

status or role in the game is derived from a functional system which also contains the other players, the chosen rules, and the various objects "at play".

For Sellars, the game of giving and asking for reasons sets up the rules for all the other games. While some might hold a strict separation between *knowing how* and *knowing that*; for instance, knowing *how* to ride a bicycle versus "knowing *that* a steady pressure by the legs of a balanced person on the pedals would result in forward motion" (Sellars 1962, 1), Sellars is comfortable treating the capacity for *knowing that* as a precondition for all instances of *knowing how*. He endorses the view that, at least for "characteristically human activity" (1), we should hold that "anything which can properly be called 'knowing how to do something' presupposes a body of knowledge *that*; or, to put it differently, knowledge of truths or facts" (1). Since only an individual "inside" the space of reasons can have knowledge of truths, the space of reasons is a kind of meta-activity which pervades and transforms all characteristically human activities (2).

Connected to Sellars' theses on reasoning and reflection are his two clashing images of man-in-the-world—the *manifest image* and the *scientific image.* The former describes the objects of common sense. These include "persons, animals, lower forms of life and 'merely material' things, like rivers and stones" (1962, 9). The latter describes the world as posited by science inclusive of the "swirls of particles, forces, and fields" of theoretical physics and "man as he appears to the biochemist, to the physiologist, to the behaviourist, to the social scientist" (20). Importantly, the scientific image is actually of *many* separate competing images in contrast to the *one* sophisticated image of common sense. But the result of this one-to-many relationship is what Sellars takes to be the central puzzle of philosophy; the scientific images are supposed to be replacements for the manifest image, but each scientific image can only replace the manifest image—and can only be related to the other scientific images—only by first presupposing the objects of the manifest image.

How are these images to be reconciled? The route taken by some phenomenologists and pragmatists is to endorse the primacy of the manifest image. This involves "interpreting the scientific image as a whole as a 'symbolic device' for coping with the world as it presents itself to us in the manifest image" (Sellars 1962, 37). The other route is to accept the final ontology taken by the scientific image at risk of reductionism. Ultimately, Sellars rejects the former instrumentalism and leans toward the latter. But he also holds that such a position will not actually compromise much of common sense so long as the individual person is viewed in relation to a

larger community. When we take up this stance, we see persons, not as mere individuals reducible to particle arrangements, but also as entities "bound up in a network of rights and duties" (39). It is this network of norms that cannot be reduced even though the objects within such a network can be.

The aim of this dissertation is to reconsider this network of norms by analyzing the space of reasons in relation to technology. I offer an *internal critique* of this Sellarsian picture against the practical backdrop of synthetic media. When one compares Sellars' position on the "clash of images" to his account of the space of reasons, his method for rescuing the manifest image seems dependent upon a rather *thin* construal of normativity, one which sees norms as arising from linguistic practice. In the contemporary landscape, this thin construal risks permitting the entry of new artificial entities into the space of reasons while obscuring the existence of non-rationalist or non-intellectualist spaces of normativity. My central question, then, concerns how the manifest image is to be reconstructed according to a conception of the human being as a technological animal, a *homo faber* as opposed to a *homo sapiens*.[2] This is a conception which takes our relation to artifacts as the primordial site of human normativity and which reconstrues epistemology as a game of giving and receiving artifactual reasons.

My attempt to reconstruct Sellars' position through a notion of artifacts is not unprecedented in the literature. Willem A. deVries (2005) holds a similar suspicion that, on the surface, Sellar does not do enough to protect common sense from reduction. To save Sellars from himself, he points to an instance where Sellars insists that the language of science be supplemented by "the dimension of practical discourse" (1965, 189) a space which is further defined as "the practical dimension of such commonsense concepts as that of what it is to be a hammer" (189, note 4). With this point, deVries shows that Sellars' reconstructed manifest image is more robust than it initially seems. According to this reading:

> A hammer is not just something with which it is possible to pound; stones are not hammers, nor are hatchets. It is not merely an empirical generalization that hammers are used for pounding. Hammers are *to be used* for pounding. There is an ought-to-be sitting in the background that is at home in a set of institutional practices constituted by still more ought-to-be's and their related ought-to-do's (deVries 2005, 275).

[2] Henri Bergson's ([1911] 1998) *Creative Evolution* provides the original model for this comparison. Hannah Arendt (1960) also provides a critical-yet-elucidatory variation of the concept. She writes that the *homo faber* "fabricates the sheer unending variety of things whose sum total constitutes human artifice" (136). This artifice, in turn, transforms the initial Heraclitean flux into a stable and objective world of things.

Though artifacts, such as hammers, are suffused with norms, there is nothing especially supernatural about these norms. Artifacts are an easy class of object to naturalize, since we can readily locate their various functions as coalescing around a common physical object. To this end, deVries holds that "even though artifacts carry along with them implicit reference to functions, to social institutions and practices, and thereby to systems of norms, they are all capable of value-free description as physical objects" (275). This is to say that the various performers in a social practice generally accept that, if the social practice were to be removed along with its instituted norms, the various features, functions, and descriptions of the constituent artifacts of that practice would find "their convergence, even their (token) identity in the particular objects" (275).

I think deVries is fundamentally correct in this interpretation of the manifest image in terms of artifacts.[3] But I also think this addendum does not let Sellars off the hook, for it still seems that Sellars' picture is one where hammers, as artifacts, only receive their normative status as objects of higher-order *linguistic practices*. But what if our engagements with artifacts were considered as a more primary kind of normativity? Consider children and some intelligent non-human animals; they too seem to live in a world of artifacts and norms without being capable of *giving reasons* for acting. The problem may also be posed in terms of a quasi-historical account. There is evidence that primitive humans developed systems of artifacts well before they possessed the symbolic resources to be reason-giving agents (Allen 2004, 2008). Are these beings normative beings after all?

In Part I of the dissertation, I argue that the manifest image is first and foremost a collection of artifacts. This view entails that it is possible for intelligent beings to follow rules, institute norms, and hold each other accountable independently of their participation in the space of reasons. In Chapter 2, I consider dueling conceptions of the normative agent as *homo faber*, the fabricating animal, versus *homo sapiens*, the rational animal. Taking inspiration from Daniel Dennett's (1987) idea of three stances—the physical stance, the design stance, and the intentional stance—I conceive of reasons as one sub-class of artifact within a broader field of artifactual norms. In my model, norms first arise through the social use and sanctioning of objects within practice. This social sanctioning yields a plurality of artifacts, which through increasing differentiation, give way to more complex forms of symbolic activity. The arrival of language institutes a different set of

[3] Amie Thomasson comes to a similar conclusion that there is no real conflict with common sense and the scientific image (2007, 148).

norms; it provides creatures with the capacity to take up the intentional stance, attribute beliefs to others, and engage in the game of giving and asking for reasons. However, since I conceive of sentences as compositions of linguistic artifacts, linguistic norms are derivative of a more basic social activity involving the use of artifacts. This *artifactual normativity* suggests a broader range of norms, inclusive of our use of both words and things. It points to the need for a model for materializing and stabilizing the manifest image as a product of our artifactual practices. I here provide such a model, which I call the space of practice.

### 1.4 – Practicing Pragmatism

Though the above sketch may look like a heavyweight metaphysical conception for how the world *really is*, my view is primarily a position about our *practices*. My view involves observations about how we navigate the social world rather than definitive claims to a true ontology. The point is to show how artifacts constitute an undertheorized system of interlocking, localized norms while also explaining how these same artifacts can be more readily integrated into our epistemic environment.

The general stance of my argument is heavily reliant on a broadly-construed *pragmatism*, one which takes the *meaning* of some object or entity to be derived from its *use*. Although this basic position was first codified in the pragmatic maxim of Charles Sanders Peirce in 1878 and reinterpreted as a theory of truth by William James, it was John Dewey who conceived of pragmatism as a general theory concerning the meaning of our tools, instruments, and artifacts (Hickman 2013, 175). I am indebted to Dewey's pragmatism most of all. My notion of artifacts, instruments, and technology is largely derivative of Dewey's own commitments and of the Deweyan reading of technology provided by Larry Hickman (1990; 2001). Yet it is important to acknowledge that pragmatist themes abound beyond Peirce, James, and Dewey.

The later works of Ludwig Wittgenstein also express a deep commitment to practice as the primary site of meaning. At the outset of his *Philosophical Investigations*, Wittgenstein ([1953] 2009) famously conceived of a series of language games of increasing complexity, the most famous of which involves builders hauling around slabs according to the vocal performances of a conducting member. It is this conception of a language game as the basic normative practice which was taken up and subsequently modified by Sellars into the space of reasons.

Though Sellars takes Wittgenstein's language games to be a position about linguistic meaning, Wittgenstein's *Investigations* is full of rich references to tools and technologies—words

function like tools in a toolbox (§11) or like the various handles of a locomotive cabin (§12). As a means of capturing the features of embodiment and practice which can be lost in these passages, Mark Coeckelbergh (2017) proposes that these references be "turned around" to show how our technologies derive their meaning from use within holistic networks just like the words and sentences of a language. In this manner, a tool in a toolbox is only a stable and meaningful use-object only when it is recognized as having a different localizable function from the rest.[4] Similarly, one knows the meaning of a locomotive handle—for instance, knowing that it is the brake handle—when one also knows what the other ones do—for instance, knowing that there are not *two* brake handles among the handles.

My own project may be seen as an extension of Coeckelbergh's. We both want to use this basic insight as a model for re-interpreting language-centered "semantic holism" into a form of "technological holism" (Coeckelbergh 2017, 3), one which takes tools, instruments, or artifacts to be basic units of meaning. To this end, despite the allure of Dewey and Wittgenstein, the most important pragmatist for the purposes of this dissertation is Richard Rorty, who along with Donald Davidson, extended semantic holism to its ultimate epistemological outcome. Both figures take Sellars' criticisms of empiricism to be a defense of the controversial epistemological position that only a belief can be a reason for another belief. The result is a view in which the world outside our bodies merely causally triggers the inner workings of belief without itself playing a justificatory role. However, Rorty further develops this contrast of reasons and causes in relation to technology in a manner that draws from the classical pragmatists and continental sources. Though technology plays a minor role in Rorty's overall project, the pluralism he exemplifies provides a useful starting point for connecting common themes in the philosophy of technology with major analytic accounts of rationality, such as those provided by Sellars, Davidson, and Brandom.

Despite his allegiance with pragmatism, Rorty often serves as a foil to contemporary pragmatist accounts since many have interpreted his positions to be a form of relativism or "linguistic idealism." As a helpful model for surveying the pragmatist landscape, Steven Levine

[4] One way of seeing the dependence relation is that there is a difference between 'using x *as* a hammer' and 'x *being* a hammer'. The Wittgensteinian point is that 'x *being* a hammer' could only make sense in relation to other classes of objects whose standard functions exclude hammering. Similarly, 'using x *as* a hammer' could only make sense given some existent x's which are hammers. And so there must be a minimal cluster of artifacts in order for each artifact to itself have clearly defined functions (and thus *be* an artifact). Importantly, this view does not moralize cases where artifacts are used outside of their intended function—there is nothing *incorrect* about using a hammer *as* a doorstop or a doorstop *as* a hammer. The point is that we cannot even explain uses of artifacts *as other artifacts* without first ascribing some norms of proper-functioning to things like hammers and doorstops.

(2019) sorts responses to Rorty in terms of two competing enterprises. On one side, there is a *communicative-theoretic approach* which—following the later work of Donald Davidson and Robert Brandom—attempts to generate a conception of language or of social practice which avoids relativism by adding certain preconditions for objectivity. On the other side, there is an *experiential-theoretic approach* which—drawing from the earlier writings of Dewey and James—attempts to ground language in the perceptual experiences of organisms (3). This second side includes classically-oriented pragmatists such as Larry Hickman (1990), David Hildebrand (2003), Mark Johnson (2007), and Thomas Alexander (2013) as well as philosophers inspired by analytic accounts of experience such as John McDowell (1994), Hilary Putnam (1995), and Richard Bernstein (2010).

Although I ultimately come to disagree with Rorty, Davidson, and Brandom, my own account is more consistent with the communicative-theoretic strand which they occupy. My position, what I call *artifactual pragmatism*, extends the communicative-theoretic account of meaning to include various non-linguistic objects as part of the space of reasons. The difference between artifactual pragmatism and experiential pragmatism is subtle. Levine writes that "the space of reasons is a space of linguistic and embodied practices, the latter being anchored in and incorporating the world through our habits and bodily skills" (2019, 230). I agree with this mixing of reason and practice. The issue is in the direction of explanation and how it redefines the space of reasons. Levine thinks that experiences are "informed by the downward influence of communicatively articulated meanings on the bodily habits and skills that organize them" (228). He intertwines reason and practice by conceiving of the space of reasons as a mechanism which reflects upon practice. I am interested in the reverse account; how bodily habits and technical skills bring non-linguistic material into our social practices of justification. I investigate the "upward influence" that the objects of practice have on the game of giving and asking for reasons. Therefore, in keeping with my intended internal critique, my account of the space of reasons retains the rationalist character of Sellars' and Brandom's formulations.

Artifactual normativity provides an account of non-rational practice which separates artifactual pragmatism from these more overt forms of rationalist pragmatism. In Chapter 2, I explore how artifactual normativity relates to the space of reasons by drawing out a tension between *homo sapiens* and *homo faber* in the work of Rorty. There I argue that our dealings with artifacts constitute a *space of practice* which is situated between the space of causes and the space

of reasons. The space of practice is where intelligent beings participate in a range of normative practices without necessarily engaging in intentional acts of justification. One is in the space of practice whenever one is engaged in an acquired activity which has socially sanctioned norms, whether this be football, bicycle riding, cooking, or piano playing. These actions are distinct from mere causal phenomena in that they are products of a social, emergent intersubjective intentionality (Allen 2008, 3).

The space of practice has two functions in my overall argument. On one hand, it locates a broader class of norms that are available for philosophical analysis. On the other hand, it performs an epistemological function; I argue that some justificatory performances utilize *artifactual reasons,* artifacts which are intentionally deployed by some interlocutor A for the purposes of belief modification in an interlocutor B. The key idea here involves seeing artifacts as related in a *rational* or *justificatory* manner to beliefs without conceiving them as sentence-like entities.

In Chapter 3, I provide an extended defense of artifactual reasons using the artifact of the photograph as a case study. Again, Rorty serves as a foil. He argues that since only a belief can justify another belief and photographs are not beliefs, then it follows that photographs cannot justify beliefs. An implication is that photographs cannot be reasons. My argument against this view consists of two steps. First, I combine Donald Davidson's account of *triangulation* and John McDowell's theory of *perceptual concepts* to provide a defense for the view that objects can be reasons. On one hand, Davidson (1990a; 1990b; 1997a) thinks that pre-linguistic social beings triangulate common objects in the world as a precondition to learning a language. On the other hand, McDowell (2008) defends a Kantian view where perceptual reasons need not be propositionally-structured entities to bypass charges of empirical givenness. Though each denies the other's thesis, together these two positions provide an account of how triangulated objects are perceptual reasons. This framework is then extended in the latter half of the chapter to show how *artifactual reasons* are unique cases where the triangulated object in question is an artifact which has already been socially indexed to do epistemic work.[5] Finally, since photographs are both objects which can be triangulated by social actors and artifacts which have been indexed to do

[5] There is a long precedent for this conception of epistemically reliable artifacts, stemming back to views on perceptual knowledge provided by D. M. Armstrong (1973) and Alvin Goldman (1976). More explicitly, Sanford Goldberg's (2020) account of *epistemic environments* has focused on instruments sanctioned by our social practices for epistemic functions. Finally, in relation to AI, Ramón Alvarado (2023) has argued for a conception of an *epistemic technology* which should be prima facie evaluated in accordance with our epistemic or rational norms rather than some set of practical or prudential norms.

epistemic work, they constitute a class of artifactual reasons which is evaluated in terms of a commonly-held epistemic reliability.

### 1.5 – Reasoning with Synthetic Media

Part II of the dissertation cashes out my artifactual pragmatism in terms of the two opening debates about synthetic media. Synthetic media results in questions concerning the status of both the "players" and the "objects at play" within the space of reasons. In relation to the "player" question, artifactual normativity is deployed as a concept which undermines the strict rationalist or intellectualist conception of normativity in present discussions of AI-reasoning, one which takes LLMs to be participants in the game of giving and asking for reasons. In relation to the "object" question, the conception of artifactual reasons is deployed to track online justificatory performances which utilize a variety of media objects to verify whether some given artifact, content, or piece of media is indeed synthetic.

My answer to the "player" question is provided in detail in Chapter 4. There I take up two competing views on the status of LLMs as participants in the space of reasons. One side argues that LLMs are legitimate reasoners as opposed to derivative instruments (Marchettoni 2025), the other side treats them as "alien" role-players which only effectively mimic rationality (Heinrichs and Knell 2021; Vallor 2024; Sambrotta 2025). Within this debate, I argue that even if current AI models were to be attributed beliefs and linguistic intentionality (in the present or near future), their lack of embodiment in an environment inhibits the basic forms of justificatory performance that are available to them. In lacking robust participation within a variety of practices, practices yielding a field of common and concrete artifacts, these systems would *inhabit* the space of reasons differently than we do. They could not, for instance, point to objects as reasons or pick up and throw world-chunks as reasons.

One potential counter to the view that LLMs lack embodiment is the existence of multi-modal LLMs (MLLMs), some of which feature both vision capabilities and supplementation with Retrieval-Augmented Generation (RAG) to reduce the probability of fabrication by linking textual outputs to "chunks" of source material (Gao et al. 2024; Huang et al 2024).[6] However, there are two problems to this counter. First, though MLLMs can "read" images, they cannot "chunk out"

[6] Most mass market LLMs are MLLMs which feature some image recognition and generation capabilities. An example of a common MLLM which features a source-grounded RAG system would be Google's NotebookLM.

image parts from inputted visual media; common MLLM models equipped with Vision Transformers (ViT)—ChatGPT, Claude, Gemini—can only analyze or compare images by first decomposing the image into token 16x16 patches and translating these tokens into corresponding textual outputs (Dosovitskiy et al. 2021). This method has limited success in practical applications like real-time image verification tasks (Chandrasekar and Jaźwińska 2025).[7]

Second, artifactual pragmatism implies a radical anti-representationalist approach to multimodal content where *images* of an object O do not possess the same content as O. This position hinges on the account of artifactual normativity. While an image of a red hammer can justify "That hammer is red" in just the same manner as a red hammer object, it cannot be *used as* a hammer in the same way—it cannot be wielded, grasped, thrown, kicked, or shoved. I argue that these non-rational uses of artifacts subsequently structure our rational expertise and our epistemic environment. If the triangulated objects indexed and used within social practices are directly taken up and utilized as reasons, then MLLMs would still lack access to this basic class of reasons even when their textual training data is supplemented by audio-visual content. The result of this line of reasoning is that disembodied LLMs are *alien reasoners* who do not genuinely participate in the space of reasons. Since these agents lack access to the embodied, non-rational practices in which other reasoners are situated, they ultimately lack access to much of the shared media which these other reasoners use as their messages.[8]

In Chapter 5, the verification of audio-visual deepfakes serves as a practical site for exploring the second question—the "object" question. Here, I take up what David Chalmers (2022) calls a limited anti-skeptical approach to deepfakes in relation to a line of social scientific research on what I call the *epistemic arms race,* a paradigm which tracks human abilities to detect AI content against technological detection methods. Against this body of research, I argue that current methods for assessing the epistemic effects of deepfakes overstate the harms by testing human detection abilities using individualistic methodology and a naïve empiricism where audio-visual

[7] Chandrasekar and Jázwińska (2025) point out that, while it is very easy for humans to check images in relation to others by merely seeing the similarities or differences, standard MLLMs often engage in circuitous or unpredictable routes in processing image contents into descriptive sub-parts. These problems are a result of machine vision methods analyzing images through pixelation patterns rather than as ordinary objects (see Alvarado 2022; Thompson and Hsu 2023).

[8] This position aligns with both Davidson's (1990c) treatment of the Turing Test and the skeptical positions on AI developed by Bender and Koller (2020) and Vallor (2024).

evidence is treated as an instantaneous perceptual "given" rather than a node within a larger web of multi-artifactual justificatory performances.[9] The result is a view which undermines human epistemic autonomy and emphasizes our reliance on AI detectors for verifying audio-visual content (Harris 2024). Against this view, I argue that in instances where online social verification of deepfakes is possible, epistemic actors engage in *artifactual deliberation.* This theory of deliberation combines pragmatist theories of political deliberation (Anderson 2006) and multi-modal argumentation theory (Groarke 1996; Roque 2015; Godden 2016, 2022) with my account of artifactual reasons in order to show how social-epistemic actors have full recourse to competing linguistic testimony, alternate confirming or disconfirming audio-visual sources, AI detectors, platform-based AI agents, and reverse-image search tools.[10] Additionally, they may use likes or other platform-based reactions as an informal form of epistemic scorekeeping, all while alternating between different tabs and therefore different arrangements of platform-specific affordances.[11]

If this conception of artifactual deliberation adequately tracks online justificatory practices, then one common framing of audio-visual synthetic media as ushering a widespread "epistemic apocalypse" is overstated (Habgood-Coote 2023). Our current social-epistemic practices of deliberation remain *robust* in a manner that prevents the total degradation of non-linguistic evidence. Instead, greater attention should be paid to the localized contexts where social deliberation among agents is inhibited, where environing artifactual reasons are minimal, or where primary harms are not reducible to epistemic or doxastic effects, for instance in the creation and proliferation of AI-generated non-consensual pornography (Burkell and Gosse 2019; Maddocks 2020).

Artifactual deliberation serves as the final social-political and methodological end which the dissertation builds. From an originary non-rational *artifactual normativity* to an account of *artifactual reasons* and finally to this modification to political practice, the positions taken up in

[9] Chronologically, the example human-centered studies I draw from are Rössler et al. (2019), Vaccari and Chadwick (2020), Korshunov and Marcel (2020), Köbis et al. (2021), Groh et al. (2022), Bray et al. (2023), Goh et al. (2024), and an overview provided by Birrer and Just (2024).

[10] Reverse-image search is treated as a useful tool for cross-referencing visual media by both Quian et al. (2023) and Weickmann and Lecheler (2024).

[11] A classic outline of affordances for social media platforms is Bucher and Helmond (2018). For an account of how platform-specific affordances mobilize disinformation, Tripodi et al. (2023) provides a set of useful case studies and Marin (2022) provides a philosophical taxonomy of effects. For an account explaining how media literacy practices of "lateral reading" effective flip between differently-organized online spaces, see Wineburg and McGrew (2019).

this dissertation are envisioned as a tentative program for identifying and mapping a multi-artifactual space of reasons. The model I propose is ultimately one which allows us to identify digital artifacts as basic objects of online epistemic environments. As a positive consequence of the model, I think that without this kind of indexing of artifacts, it would be tedious, if not cognitively impossible, to track the broad range of available justificatory performances that we utilize in everyday life. When properly mapped, the artifacts taken up within these performances not only justify our *beliefs* but also stand in justifying relations to *other artifacts* in an admixture of practical affordances and epistemic configurations. It is these configurations which must be further explored if artifactual pragmatism is to truly be applied as a theoretical device.

And so, my project concludes as but another messy and incomplete picture of communicative practice in the twenty-first century. But it is a much needed one. It is a picture which trades the thin, abstract formalism of Sellars' *logical* space of reasons for the thick, concrete infrastructure of an *artifactual* space of reasons—one where photographs and thermometers, rocks and rivers, slabs and sentences are all objects at play in that great meta-game we call reason-giving.

# PART I:

# REASONS AND ARTIFACTS
# IN CONTEMPORARY PRAGMATISM

# CHAPTER 2

## ARTIFACTUAL PRAGMATISM:
## ON REASON AND PRACTICE AFTER RORTY

> One of the best things Bergson ever said is that homo sapiens is homo faber.
>
> — John Dewey, "Letter to John D. Graves" (July 27, 1949)

### 2.1 – Introduction

Though adopting an analytic variant of pragmatism which focused extensively on linguistic practice, Richard Rorty unexpectedly maintained an ambivalent relationship to the linguistic turn throughout his career. While his early anthology, *The Linguistic Turn,* earned him the reputation of being an adherent of the analytic philosophy of language running from Bertrand Russell to P. F. Strawson, Rorty's introduction to the collection already hinted at an underlying suspicion toward philosophical problems, concluding with an ironic proclamation that the works he had collected would eventually encounter "the pattern of creeping obsolescence" (1967, 39). Rorty's general dismissive attitude toward the linguistic turn was more explicitly revealed over time. In *Philosophy and the Mirror of Nature*, he denies the common narrative, attributed to Hilary Putnam, that the linguistic turn "enabled philosophers to provide substantive solutions to traditional problems" (Rorty 1979, 258, n2). And in his later work, even after flirting with elements of literary theory, Rorty was still keen to reject any sort of linguistic methodology which analyzes concepts or which isolates language as an identifiable structure over and against the rest of the world. In his final set of philosophical papers from the early 2000s, he firmly reiterated that "there is no interesting sense in which philosophical problems are problems of language" (Rorty 2007b, 160) and followed Donald Davidson in rejecting an identifiable distinction between "knowing a language and knowing our way around the world generally" (Rorty 2007a, 159).

Nonetheless, any reader of Rorty cannot help but notice that his works are densely populated with references to language, linguistic practice, sentences, and propositional attitudes. Rorty's status within philosophy and pragmatism-writ-large is taken to be the extreme position that

there is no escaping or transcending linguistic practice, even by referring back to states of direct experience.[12]

In this chapter, I want to complicate the standard reading. I do not intend to argue against the common reader or the philosopher turned quasi-historian by challenging Rorty's overt linguisticism, but I do want to draw out an underlying thread in Rorty's work—his analysis of artifacts.

The starting point for such an analysis is what Robert Brandom, has identified as a "language-as-tool metaphor" (Brandom 2011, 77–78), a device which Rorty often uses to deflate or undermine the representational status of language. The metaphor is effectively deployed to reject the status of language as a medium (1989, 13), to reduce beliefs to "habits of action" (1991c, 93), and to blur the distinction between "texts and lumps" (1985, 89). Yet the critical force of the position always seems lost; Rorty always seems to revert back to some more standard account of linguistic meaning, some more literal account of belief, or some more standard separation of the lumpy world from the logical space of reasons. These tensions serve as the grounding themes of this dissertation.

In the first part of this chapter (Section 2.2), I draw out the problems which emerge from Rorty's inability to reconcile the "internal" standpoint of beliefs and the "external" standpoint of our coping with the world as a tool-using animal. This tension results in both i) a thin conception of normativity as resulting entirely from attributions of belief and ii) what many take to be Rorty's most radical thesis, that the world outside our system of beliefs is not properly organized to justify this system of belief.

In the second part of the chapter (Section 2.3), I show how the tension may be recognized as a clash between two conceptions of the normative subject—the rational *homo sapiens* and the technical *homo faber*—each taken up and defended by one of Rorty's students. On one side, Robert Brandom (1994) extends the internal picture of beliefs outward to reconstruct the world into states of affairs capable of, themselves, being reasons for beliefs. Alternatively, Barry Allen (2004) extends the external picture of tool-use inward in order to undermine the propositional bias

[12] For various accounts of Rorty's standing in pragmatism, see Cheryl Misak (2014), Susan Haack (2004), Richard J. Berstein (2010), and Steven Levine (2019). For an in-depth comparison of Rorty to the classical pragmatists, see Chris Voparil (2021).

inherent in Rorty's understanding of the epistemic subject. Both strategies have their positive and negative aspects, but neither fully reconciles the tension.

To resolve this stand-off, I develop from the three above positions an account I call *artifactual pragmatism* (Sections 2.4–2.6). My account, at its core, decomposes our social practices into relations of artifacts. Artifactual pragmatism consists of two novel claims. First, it widens the gap between reasons and causes to permit an intermediary *space of practice.* This space accommodates what I call *artifactual normativity,* the notion that non-linguistic practices are object-related and world-involving such that artifacts are basic conceptual and normative entities. As a second position, artifactual pragmatism argues for a conception of *artifactual reasons*, where our artifactual practices are identified as the mechanism through which the lumpy matter of the world can be brought into our practices of giving and asking for reasons.

### 2.2 – Rorty's Linguistic Turn

In the wake of John Dewey, the remnants of classical pragmatism found refuge as an undercurrent in the burgeoning American analytic philosophers of the mid-twentieth century. Both Willard van Orman Quine (1951, 46) and Wilfrid Sellars (1950, 314) describe their outlooks as particular variants of pragmatism, yet despite these overt declarations, it was the publication of Richard Rorty's *Philosophy and the Mirror of Nature* in 1979 which is credited with reviving pragmatism within the American academy. Rorty's master argument takes the form of a comparison wherein Quine's empiricist pragmatism is set in opposition to Sellars' rationalist pragmatism as a means to rid post-Deweyan pragmatisms of foundationalism. While it is doubtful, if not outright dismissed, that Rorty's master-argument captures a return to Deweyan pragmatism, the basic positions Rorty adopts through the Quine-Sellars synthesis forms the basis of his general understanding of how pragmatism can be derived within analytic philosophy, and therefore, they are the starting point for my own analysis.[13]

[13] Rorty is clearly indebted to Dewey's analysis of "immediate knowledge" within the later *Logic* — "The two doctrines, that there is an immediate knowledge of existential objects or of qualities as sense-data, and that there is an immediate knowledge of rational principles—necessarily go together. Atomistic empiricism and rational a priorism are correlative doctrines" (LTI, LW 12: 156).

*2.2.1 – The Quine-Sellars Synthesis*

The Quinean development utilized within Rorty's argument involves aligning pragmatism with a more thoroughgoing empiricism. Quine concludes his famous "Two Dogmas of Empiricism" by noting the "pragmatic stand" taken by C. I. Lewis and Rudolph Carnap regarding the choice of ontology; he holds that "their pragmatism leaves off at the imagined boundary between the analytic and the synthetic" (1951, 46). The Lewis-Carnap position supposes that analytic statements can be analyzed purely on the basis of their meaning and therefore introduces a specific philosophical methodology of conceptual analysis which can be identified apart from science. Quine thinks that he is undermining this distinct methodology by denying that there is a suitable treatment for analytic statements which does not fall into circularity. Without this special class of statements, the analysis of concepts cannot be distinguished from empirical science more broadly. The "conceptual analyst" of the ivory tower amounts to nothing more than a second-hand field linguist, and "necessary truths" are nothing more than solidified beliefs that are harder to refute without demolishing the entire doxastic web.

Yet Quine's initial program falls into an empiricist trap when he aligns himself with a certain phenomenalist strand of the logical positivists, who assert subjective experience as the primordial content of any and all conceptual schemes. He therefore holds that the physical world is organized from an initial stream of experience, that "within the phenomenalistic conceptual scheme, the ontologies of physical objects and mathematical objects are myths" (Quine 1948, 19). And later states that the myth of physical objects is only superior to the gods of Homer because "it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience" (Quine 1951, 44).

In contrast to Quine, Sellars' program is focused on reorienting the relationship between physical objects and experience such that the former is epistemologically and ontologically prior to the latter. His famous "Empiricism and the Philosophy of Mind" is a long and complex dismantling of the idea that immediately "given" contents of experience serve as a foundation for knowledge. The basic picture of Sellars' ([1956] 1997) account is that perceptual experience, alone, cannot provide *premises* on which empirical knowledge is established. This results from an inconsistent triad consisting of the following positions: 1) that to *sense* something is red is to *know* that it is red, 2) that the ability to sense sense-contents such as red is not acquired, and 3) that the ability to know facts in the form "X is red" is acquired ([1956] 1997, 21). To endorse all three

positions in this inconsistent triad is to endorse the *Myth of the Given*, where nonconceptual and immediately experienced sensory states serve as non-inferential knowledge or as premises to further kinds of inferences despite not being of the same class or kind as the rest of the edifice of knowledge (33).

Sellars' solution to this dilemma is to posit a radical view he calls *psychological nominalism*, where "*all* awareness of *sorts, resemblances, facts*, etc., in short, all awareness of abstract entities – indeed, all awareness even of particulars – is a linguistic affair" ([1956] 1997, 63). This view amounts to claiming that knowledge *that something is red*, must presuppose the ability to place red sensations into a single class or concept and also differentially relate this concept to others. For example, one cannot know that "The cube is red" without knowing the concepts 'cube' and 'red', and knowing these concepts requires minimally knowing other concepts like 'sphere' and 'green'. In this manner, a being which sees red, but does not have the linguistic capacities to sort this episode into concepts would not count as a knower—the seeing red is not alone sufficient to know that 'Some x is red'. Or as Sellars puts it in an earlier essay:

> as children we learn to understand the noise "blue" in much the same way as the dog learns to understand the noise "bone," but we leave the dog behind in that the noise "blue" also comes to function for us in a system of rule-regulated symbol activity, and it is a word, a linguistic fact, a rule-regulated symbol only in so far as it-functions in this linguistic system (1950, 302).

The issue with Sellars' denial of privileged experience is that this itself leads him to assert the methodological priority of conceptual analysis. Insofar as language is "rule-regulated" it is "essentially an axiomatic system" (310). This set of rules then provides a mechanism for analyzing language as a formal system. And so, in thinking of experience as already organized by means of a rule-governed system, Sellars denies givenness only to embrace some form of conceptual analysis which still treats of necessary truths.[14]

Rorty's use of both Quine and Sellars amounts to undermining Quine's talk of "experience" using Sellars's criticism of the Myth of the Given and denying Sellars' talk of necessity by using

[14] Patricia Churchland provides an inside look at the contrast between Sellars and Quine. Recounting her time reading *Word and Object* at the University of Pittsburgh in the 1960s, she notes how Sellarsian students in a seminar room sided with the anti-Quinean faction "who adhered to idea that conceptual analysis revealed necessary truths about the way things are and the way the mind works" (Churchland 2013, xi). The Quinean faction insisted that language was a physical phenomenon that could not produce any conceptual truths and that all statements were open to revision. Though the Sellarsians protested with thought experiments, the Quinean's treated these as "parochial, circular, or uninformed" (2013, xii).

Quine's criticism of the analytic-synthetic distinction. Rorty locates the target of Quine's critique as the rationalist philosopher "who knew about concepts by knowing about the meanings of words, and whose work therefore transcended the empirical" (Rorty 1979, 272). Sellars' target is the philosopher who grounds knowledge of the common world in some directly experienced content. Together, Rorty thinks that Quine and Sellars undermine the philosopher's privileged access to something beyond common sense, some neutral standard which would allow the philosopher to arbitrate disputes at a "deeper" level than anyone else.[15]

In denying a privileged position on which philosophy could rest, Rorty's alternative is a position, he calls "epistemological behaviorism" which asserts that "philosophy will have no more to offer than common sense (supplemented by biology, history, etc.) about knowledge and truth" (1979, 176).[16] The philosophical method suggested by epistemological behaviorism is but a "useful kibitzing" where philosophers engage in intellectual dialogue using their extensive background with historical and contemporary theoretical debates as well with commonly-held arguments (393). The epistemic standpoint of the philosopher is justified, not by their knowledge about the workings of knowledge, but by their historical involvement within a longstanding intellectual tradition and their practical expertise in constructing and demolishing arguments.[17]

Rorty modifies this initial position in his later work, moving to a view he calls *anti-representationalism*. Anti-representationalism holds that knowledge is not "a matter of getting reality right," but instead "a matter of acquiring habits of action for coping with reality" (1991b, 1). Rorty expresses that the purpose of this viewpoint is to set aside realism-antirealism debates within philosophy. Once one treats human beings as just another kind of animal and once one understands that language arose as a natural phenomenon, then both the skeptical thesis that

[15] For an extended treatment of Rorty's combined use of Quine and Sellars, see Koopman (2011, 65–68).

[16] There are several aspects about epistemological behaviorism which are important to note. First, Rorty admits that it is a name for what he would otherwise call pragmatism (1979, 176). Second, epistemological behaviorism is the culmination of a thread that runs underneath *Philosophy and the Mirror of Nature* where Rorty suggests that a loose, holism is compatible with and constitutes common sense, and that all attempts to avoid this holism are attempts to dig underneath common sense and to collapse the manifest image into some alternative ontology. Third, Rorty's treatment of common sense is notably not a transcendental argument in line with Strawson's treatment of material objects as necessarily required for experience (see Strawson 1966). Instead, he believes that transcendental arguments of this kind reduce to the ontologically agnostic argument that *beliefs* about experience must presuppose *beliefs* about objects (see Rorty 1970, 248; Rorty 1971, 262–263).

[17] There is a certain kind of intellectual conservatism in Rorty's treatment of philosophy as an intellectual tradition. In particular, Rorty thinks the philosopher is caught up in what conservative political philosopher Michael Oakeshott terms "the conversation of mankind" (1979, 389).

humans are fundamentally out of touch with the world beyond their skins and the realist thesis that humans need to establish word-world semantic relations to secure their access to reality become unnecessary. As a result, Rorty thinks there is no need to figure out how to unite all theoretical descriptions into "one grand super vocabulary" (1989, 11). The philosopher should "treat alternative vocabularies as more like alternative tools than like bits of a jigsaw puzzle" (1989, 11). The "grand super vocabulary" of old can be reconstructed only as a toolbox. All vocabularies fit together once one sees how humans, as *language using* animals, use different descriptions as better-or-worse instruments for achieving identifiable aims.[18]

### *2.2.2 – Tools and Texts*

While Rorty's general program endorses philosophical humility and frames philosophy as but one part of an interdisciplinary conversation, the details of Rorty's account produce tensions that are not easily resolved. On one hand, Rorty's conception of the normative subject seems based around the possession of linguistic concepts and the capacity to engage in conversation. But on the other hand, Rorty consistently uses the language-as-tool metaphor to blur this linguistic demarcation by providing a quasi-historical account of how linguistic practices arise from the creation of particular strings of linguistic artifacts. What Rorty wants is both a Sellarsian conception of linguistic normativity and a Quinean redescription of linguistic rules in terms of behavior—but these two standpoints are at odds with each other in their conceptions of normativity.

The tension is present but less pronounced in the work of Donald Davidson, with whom Rorty often aligns. Davidson too takes up both lessons from Quine and Sellars. Against Quine, Davidson rejects the empiricist notion that there is some experiential content which must be organized by a conceptual scheme, preferring a casual holism where the ordinary objects of the world are the entities organized within language without the existence of perceptual intermediaries (Davidson 1973).[19] Against Sellars, Davidson holds that there are no fixed linguistic rules, that the

[18] Against Sellars, Rorty appears comfortable with treating the scientific image as a symbolic instrument to be used within the manifest image. As a pluralist, he thinks that all "philosophical images," inclusive of alternative metaphysical and political theories, can be united by treating them in such a manner. His endorsement of what Arthur Fine calls "the natural ontological attitude" speaks to the notion that the irreducible common-sense ontology needs no supplementation except by more-or-less useful linguistic instruments, instruments which can themselves be reduced to the common-sense ontology (see Rorty 2004).

[19] Davidson first challenges the notion that language organizes a single, all-encompassing world: "We cannot attach a clear meaning to the notion of organizing a single object (the world, nature etc.) unless that object is understood to contain or consist in other objects. Someone who sets out to organize a closet arranges the things in it. If you are told

only way to make sense of the shared human capacity to understand communicative phenomena such as idioms, metaphors, malapropisms, or the multi-lingual puns of *Finnegan's Wake* is to hold that we always assume the standpoint of radical interpretation (Davidson 1978; Davidson 1986). The resulting position is a loosely anthropological externalism where perception, action, and justification are understood from a third person perspective of the subject and where no mental or perceptual intermediaries are posited between the subject and the world (Davidson 1990a).

Rorty explicitly endorses Davidson's positions, leading them to hold almost identical views throughout their careers. But where the Davidson-Rorty alliance comes apart is on the acceptability of the language-as-tool metaphor as a device for naturalizing intentionality. Davidson holds that "language aptitude is part of our natural equipment, and not a tool we contrived for coping with problems of understanding, calculation, and communication" (Davidson 1997, 133). In contrast, Rorty sees no issue comparing language to tools. He writes that "language is a set of tools rather than a set of representations – tools which, because of what Dewey called 'the means-end continuum', change their users and the products of their use" (Rorty 1991a, 3). He then uses the language-as-tool metaphor to explain the attribution of intentional states to organisms; beliefs and desires are posited as sentential attitudes by our use of "strings of marks and noises" as tools to predict the behavior of others (5). Rorty holds that if beliefs are interpreted as "habits of action" asserted from outside of the self, then the mind-body problem dissolves into the distinction between "the organism's movements and the interior states of the organism which you need to posit in order to explain and predict those movements" (1991c, 93).

Since beliefs are posits produced by our use of "marks and noises" to predict an organism's behavior, one might wonder whether Rorty's position has evacuated beliefs of their assumed intrinsic intentionality. As Rorty sees it, this worry is deflated so long as "marks and noises" are products of an irreducible, shared vocabulary. So long as we keep speaking to each other, we will keep attributing intentional states to each other. But this resolution generates the *thin* notion of normativity that I identified in the introduction. It appears as though human behavior is to be understood, in behavioristic terms, as causal responsiveness to stimuli *until* beliefs are posited

not to organize the shoes and shirts, but the closet itself, you would be bewildered" (Davidson 1973, 192). He then extends the criticism to sense-datum theories: "For how could something count as language that organized *only* experiences, sensations, irritations or sense data? Surely knives and forks, railroads, mountains, cabbages and kingdoms also need organizing" (1973, 192).

within a community to explain this behavior—only then can this behavior be characterized as following norms. However, I will show below that this choice of explanation is unmotivated.

The adoption of an external perspective for describing a creature's behavior provides a robust standpoint for explaining normativity once linguistic behavior is understood to be one of many forms of identifiable behavior. The issue for Rorty and for Davidson is that they take up an externalist perspective while also holding fast to Sellars' psychological nominalism, that all awareness is a linguistic affair. This leads them to track, from the external perspective, only those performances which they deem necessary for beliefs, as sentential explanations of behavior, to be instituted. Rorty's instrumentalism, then, adds insult to injury by treating beliefs as the effect of some initially predictive and operational activity by pre-normative beings.

And so Rorty's explanation for the emergence of normativity is an instrumentalist variation of Sellars' linguistic normativity, one which makes use of Davidson's externalism and Dewey's earlier conception of language as the "tool of tools" (EN, LW 1: 134). But the loose anthropology that follows from Rorty's epistemological behaviorism makes this choice an uneasy one. As I will show below through the work of Barry Allen, Rorty could just as well understand marks and noises as linguistic *artifacts* within a broader repertoire of artifactual performances. This alternative choice makes room for beliefs while also generating a thick conception of normativity where object-related localized norms emerge for each and every human artifact.

### *2.2.3 – Coherentism, Linguistic Idealism, and External Friction*

The thin conception of normativity defended by Rorty generates a second problem related to epistemic justification. Though Rorty treats beliefs as "habits of action" which are expressible in terms of sentences, he also maintains a largely internalist conception of epistemic justification where only a belief can justify another belief.

Here, Rorty again joins Davidson, who neatly separates a person's beliefs from the external causal forces outside of this system. In order to avoid falling into the Myth of the Given, Davidson adopts the view that only a belief can justify another belief (Davidson 1983). He defends this view using the Sellarsian argument concerning the inability to logically or rationally relate sensations to beliefs:

> The relation between a sensation and a belief cannot be logical, since sensations are not beliefs or other propositional attitudes. What then is the relation? The answer is, I think, obvious: the relation is causal. Sensations cause some beliefs and in this sense are the basis

> or ground of those beliefs. But a causal explanation of a belief does not show how or why the belief is justified. (Davidson 1983, 162)

Rorty radicalizes this position. If only entities like sentences, and by extension sentential attitudes, can be true, then truth is not a feature of the world outside of human practice. He writes, "Truth cannot be out there – cannot exist independently of the human mind – because sentences cannot so exist, or be out there" (1989, 5). The result is an epistemological standpoint where the epistemic subject cannot transcend or escape their own system of beliefs. It is a coherentism where new perceptual states are translated into beliefs and then subsequently integrated into an already-self-standing edifice of other beliefs.

To avoid accusations of "linguistic idealism," that the world is linguistically constructed all the way down. Rorty admits that the world does causally impinge upon our system of beliefs. He adds, "To say that the world is out there, that it is not our creation, is to say, with common sense, that most things in space and time are the effects of causes which do not include human mental states" (1989, 5). He therefore holds that "there is such a thing as brute physical resistance – the pressure of light waves on Galileo's eyeball, or of the stone on Dr. Johnson's boot" (Rorty 1985, 81). But he rejects translating these causal relations into justificatory relations. The world can only be understood to push back upon the *entire system* of beliefs. Therefore, nothing like facts, evidence, or even material objects is given immediately to perception.[20] The role once played by objectivity —the notion that the singular subject is directly confronted with reality—can now only be captured by *solidarity*—the collective conversation of a variety of creatures passing around different justifications for their beliefs (see Rorty 1984). Truth becomes something to steer our practice by (see Rorty 1996).

This position clashes with Rorty's earlier understanding of beliefs as "habits of action" which interact with the external environment. While the externalist position sees persons interacting with the commonsense world, Rorty now admits that "'recontextualizing objects' could just as well be called 'recontextualizing beliefs'" (1991c, 101). In this manner, the external standpoint is just a coherent *belief* Rorty holds among his internal web of beliefs. Again, Rorty wavers from the most radical conclusion by reminding the reader that, if language emerges as a

[20] Both Rorty and Davidson have the penchant for restating this position in skeptical terms where "the world" plays no epistemic role in belief formation. For example, Rorty's (1972) "The World Well Lost" and Davidson (1973). Though Davidson thinks "the world" is to be replaced with "unmediated touch with the familiar objects whose antics make our sentences and opinions true or false" (1973, 198).

natural phenomenon, there is no way to be radically out of touch with the world. The skeptic, in contrast to Rorty and Davidson, assumes that "our language-game may have nothing to do with the way the world is" but Rorty thinks this is "a picture of a gear-wheel so out of touch with the rest of the mechanism as to be spinning idly" (1986, 141). He thinks this picture only arises when we believe in an anti-naturalistic picture of the subject, one which is not justified by our current lights.

Despite Rorty's rejection of skepticism, in *Mind and World,* John McDowell famously accuses the Rorty-Davison position of degenerating into a "frictionless spinning in a void" (1994, 11). Though McDowell equally rejects the mythic, empirical Given, which he redescribes as the idea that "the space of justifications or warrants, extends more widely than the conceptual sphere" (1994, 7), he thinks that the "confinement imagery" which Rorty and Davidson advance produces an equally untenable picture which prevents thought from being in touch with the world (15–16). To rescue a basic conception of objectivity, that the world constrains thought in a justificatory, rather than merely a causal manner, McDowell thinks it is necessary to provide an account of conceptuality which is not limited to linguistic items or the contents of belief. Otherwise, Rorty and Davidson lack the "external friction" which the world is supposed to impose upon our thinking (11).

### 2.3 – Pragmatism After Rorty: Homo Sapiens or Homo Faber?

To trace the narrative as I have presented it so far, Richard Rorty's Quine-Sellars synthesis runs into two interrelated problems. First, the language-as-tool metaphor threatens the notion of intentionality and normativity that Rorty adopts from Sellars. Second, the coherentist picture that only a belief can justify another belief threatens to undermine the naturalism and anti-skepticism that Rorty endorses.

If one treats Rorty as fundamentally endorsing naturalist and externalist viewpoints, then beliefs are treated as effects of a natural creature's linguistic practice, but this conception seems to produce a thin and explanatorily weak conception of our norms. One might ask why marks and noises have been chosen as instruments for naturalizing intentionality if their use coincides with the use of a variety of other instruments and artifacts.

On the other hand, if one treats Rorty as fundamentally an internalist and a coherentist, then the world beyond our skins cannot play a justificatory role in our affairs. One might ask how

the external standpoint in which language is asserted to be a tool can even be related to the "internal" web of belief without itself being converted into a belief. Is Rorty's starting point the tool or the text?

This tension is taken up and advanced in different directions by two of Rorty's students within the volume *Rorty and His Critics*. Taking up a Sellarsian stance, Robert Brandom re-entrenches language as the central object of philosophical inquiry, pivoting towards rationalism and recasting the world as states of affairs capable of justifying belief. Alternatively, Barry Allen takes up a more anthropological approach that appeals to our broader use of tools in order to undermine Rorty's understanding of knowledge as justified belief. Each makes some novel move in relation to Rorty by distilling and defending some variant of the normative subject which was implicit in Rorty's work, either the rational *homo sapiens* or the technical *homo faber*. Yet neither manages to show how normativity operates outside of language or how beliefs relate to the world; Brandom ultimately trades non-linguistic normativity and the materiality of the world for a form of idealism and Allen can only satisfy the demands of non-linguistic normativity and external friction by defending a revisionist conception of knowledge.

### *2.3.1 – Brandom's Rationalist Pragmatism*

In "Vocabularies of Pragmatism," Robert Brandom (2000c) challenges several of Rorty's key commitments, three of which I will address here: (1) his rejection of language-independent truths, (2) his restrictively internalist picture of epistemic justification, and (3) his instrumentalism about language use. For each of these three challenges, Brandom relies upon a strict separation of causation and justification which attributes separate laws to each domain. As he later reconstructs it, his position draws a sharp line between *practical intentionality* referring to "the sort of skillful practical coping nonlinguistic organisms exhibit" (2019, 85) and *discursive intentionality* as the ability to justify ourselves through linguistic performances (2019, 82). The core of Brandom's view is that both Rorty and the classical pragmatists make a grave error in collapsing the norms of discursive intentionality into practical intentionality rather than providing an account of how discursive intentionality is specifically tied to the normatively autonomous practice of reason-giving. Although he takes up the pragmatist dictum that meaning comes from use, Brandom differentiates himself from past pragmatisms by locating his position relative to rationalism and

anti-naturalism.[21] As such, he defends the view that the correctness or incorrectness of our linguistic performances cannot be reduced to natural phenomena, that such logical notions as "contradiction, correct inference, correct judgment" are not barred by the laws of nature, but rather by mutual sanctioning of linguistic rules in a social practice (1994, 12). This account ultimately accepts most of Rorty's premises while denying his conclusions. Human beings start as natural creatures that use marks and noises to get what they want. However, at a particular level of complexity, these marks and noises express a logical structure to which human beings—now as *rational beings*—must submit.

The pivot point of this transformation is the capacity for making inferences. Brandom's position is a *semantic inferentialism* which makes inference the key concept for possession of discursive intentionality and for grasping the meaning of a concept. In *Making It Explicit*, Brandom (1994) famously frames the centrality of inference in relation to the vocal behavior of a parrot. While parrots can reliably respond to external stimuli with complex vocal performances, for example, by stating "That's red" in the presence of red objects, they cannot make inferences between linguistic concepts of similar and different concept classes. They do not know that "That's red" is an assertion which is incompatible with "That's green" or compatible with "That's colored" (1994, 89). As a result, they do not truly understand the meaning of the term 'red' as a component of a proposition and cannot be held responsible for error. This latter point is crucial for understanding Brandom's normative point—while rational beings can be held responsible for lying, we cannot tell whether a parrot who says "That's green" in the presence of red objects is intentionally acting out as opposed to merely unintentionally misapplying the words. The same goes for "artificial instruments" such as thermometers or spectrometers which are socially sanctioned to respond correctly to some class of stimuli (88). While we rational beings understand when instruments make an error, the instruments are not themselves responsible for this error, nor can they understand themselves as making an error.

Nonetheless, while Brandom thinks neither parrots nor thermometers are rational, he does not make the same mistake as some early modern rationalists in denying feeling to animals.

[21] Broadly, Brandom endorses a reasons-first account of normativity which argues that only those agents who can play in the game of giving-and-asking-for-reasons are agents who can be held morally responsible. Since, this game is an emergent social practice, Brandom does not think that normative notions of rightness, correctness, or the good map to an organism's experiences of pain and pleasure. He would claim that norms are only made explicit by full-blown language users.

Parrots, unlike thermometers, possess *sentience* in the sense that they are aware and awake (1994, 5). But the parrot does not possess *sapience* in the sense that "one explains its behavior by attributing to it intentional states such as belief and desire as constituting reasons for that behavior" (5).[22] Sapient beings are those that possess *discursive intentionality*, who follow rules, and who are concept-mongering (8). In possessing beliefs and treating others as possessing beliefs, they engage in what Brandom calls *deontic scorekeeping* by keeping track of each other's commitments and entitlements (1994, 142). Essentially, sapient beings use language to attribute internalized claims to others (in the form of beliefs) and assess these beliefs according to whether the believer is justified. The totality of scorekeeping among all sapient beings constitutes the space of reasons, where linguistic reasons for and against beliefs are assessed and the results of different justificatory performances on each interlocutor's beliefs is tracked by others.

With this extensive edifice in place, Brandom (2000c) first targets Rorty's criticism of representationalism. Once we see the space of reasons as an autonomous practice of scorekeeping above-and-beyond other practices, we can see how linguistic concepts are fixed by practices in a manner that makes objectivity possible. While Rorty treats all norms as determined by social practice, Brandom thinks that social practice is a requirement but not a criterion for correctness. Some act of practical intentionality like a greeting gesture or a ceremonial rite is entirely subject to the internal norms of a group. In these cases, the group, and not the world, determines whether some performance is correct or incorrect. But Brandom holds that linguistic performances are of a kind such that the linguistic community could be collectively wrong in their beliefs (1994, 53). A statement like "The molecular formula of water is $H_2O$" requires language to fix the meanings of terms, but once this meaning is fixed, the world determines the statement's truth, not the linguistic community.

Rorty is therefore correct in his starting point, that representation is not a mental activity nor a theory for how language works, but Brandom's point is that this does not preclude representation being a social achievement of language users. Rather than first representing objects and then inferring from these representations, Brandom holds that our capacity for inference

[22] One might get around the difficulties of Brandom's treatment toward animals by connecting the sentience-sapience distinction to that of moral patients versus moral agents (see Tom Regan 1983). Under this reading, Brandom is not denying that animals are *moral patients* who should be treated with respect. Rather he is denying that animals are *moral agents* capable of intentionally committing acts of wrongdoing. He might then say that as rational beings we are entitled to treat animals well, even though as non-rational creatures they are not entitled to treat each other as such.

transforms our reliable responses to each other's vocal phenomena into a logical edifice through which accurate representations are made possible. Once a language is set up, our *claimings* about the world can be mapped to some *claimed* thing (2000c, 162). But crucially, Brandom holds that we can also introduce facts as "claimables" or things which would make possible claims true (162). With these distinctions, Brandom agrees with Rorty that "there were no true claimings before there were vocabularies" but this does not mean that "there were no true claimables" (162). The potential for some content to be claimed presupposes that this content is already structured in a logical or conceptual manner prior to any claims. As Ryan Simonelli (2021) helpfully elucidates Brandom's position, Brandom adopts Sellars' psychological nominalism but rejects his ontological nominalism. The result is that Brandom treats the world in-itself as a linguistic affair.

Brandom (2000c) carries this point about claimables to Rorty's contention that only a belief can justify another belief. First, he asserts from the existence of claimables that "there are facts, that is, conceptually structured truth-makers" (164). With a second point about how these facts, by definition, relate in a justificatory—rather than a strictly causal sense—to beliefs, Brandom argues that "in a central range of favored cases of perceptual experience, the facts *are* the reasons that entitle perceivers to their empirical beliefs" (164). In other words, if each claim can be matched to a perceived claimed or a claimable entity, we should hold that the claimed or claimable entity is what makes a claim true, not social solidarity. As a third move, Brandom (2000c) connects this point to an anti-relativist, externalist position on knowledge (165–166). He argues that persons who reliably respond to the facts can count as knowers even when they are unaware that this relationship holds. In Brandom's view, person A counts as a knower so long as some other person B is in a standpoint to assess the relationship between A's beliefs and the facts (166). The result of this position is quite radical. Accordingly, we are in a position, as B, to say that Galileo as A, knew for a fact that "The Earth revolves around the Sun", even though the social practices of Galileo's time would have rendered his claim unjustified.

All of this then bears upon Rorty's instrumentalism. If linguistic performances can attain representational correctness, then they are more than instruments for getting what we want.[23] First of all, this is because claims can be *about* the world in a way that hammers and wrenches cannot

[23] Brandom frames the problem of instrumentalism as one of objectivity in a manner not unlike Bertrand Russell's earlier criticism of James' theory of truth in Russell ([1909] 2005). Within critical theory, Max Horkheimer (1947) also problematizes the pragmatists' defense of a "subjective reason" that is solely concerned with means and ends (1947, 3), which he sees exemplified in Dewey's account of truth (42).

be. Second, they belong to a class of entities which are evaluable according to the logical standards of the space of reasons rather than in relation to the instrumental end of some local practice. If Brandom is correct, then Rorty cannot fully dispense with representationalism and objectivity. Instead, he must hold that representation and objectivity emerge from non-representational, intra-linguistic rules. Rorty can assert that there are no perceptual intermediaries to knowledge, no givens, and no a priori reasons, but he is wrong to think that these premises render objectivity impossible.

*2.3.2 – Allen on Artifactual Performance*

In "What was Epistemology?" Allen takes up a very different strategy. By appealing to our use of artifacts as normative performances, Allen blocks Brandom's appeal to a linguistic normativity and circumvents the need for representation by grounding our practices in a world of material objects. As a central move, Allen points out that Rorty often resorts to a false dichotomy between private, representational mental events on one hand, and public discourse on the other (Allen 2000, 224–225). If one follows this line of thought, then the only means of undermining representationalism is to embrace a loose conversationalism, where deliberative consensus determines the normative correctness of conduct and the truth-values of sentences. Like Brandom, Allen finds this consensus-oriented account normatively problematic, but unlike Brandom, he wants to preserve Rorty's anti-representationalism and avoid falling prey to talk of facts and states-of-affairs.

Allen (2004) instead considers a constraint on knowledge which is neither representational nor consensus-driven. He asks, "Why this choice? Couldn't the relation between language and the nonlinguistic be like that of a tool: not a representation, and not a consensus, but an effect and artifact?" (2004, 157). For his failure to see this alternative, Allen (2000) accuses Rorty of holding several "biases of epistemology," including a *propositional bias*, a *discursive bias*, and a presumption of knowledge as "belief-*plus*" (2000, 226–227). Together these positions articulate a view of knowledge which exclusively considers knowledge-that (over knowledge-how) and which sees this kind of knowledge expressed only in claims or in beliefs.[24]

[24] Many of the biases Allen identifies are also present in the "thing knowledge" of Davis Baird (2004). Baird equally criticizes what Allen calls the "belief-*plus*" account: "Instruments, whatever they may be, are not beliefs" (4). He subsequently connects this to a "text bias" which obscures the material aspect of knowledge production (5).

While Brandom wants to avoid Rorty's pitfalls by reconstructing pragmatism as a semantic thesis with epistemological implications, Allen implies that Rorty is too wedded to the standard epistemological vocabulary to truly dispense with the problems of philosophy. If Rorty really wanted to get rid of standard epistemological problems, he would have taken up the *artifact*, not the proposition, as the basic unit of knowledge. In doing so, he would have seen that propositions are linguistic artifacts which cannot be disentangled from the broader system of non-linguistic artifacts in which they arise, or as Allen puts it, that propositional and discursive knowledge can only be enacted "because of more primary interactions with more primary (material) artifacts" (Allen 2004, 17).

Allen's initial criticism of Rorty sets the stage for his alternative thesis of knowledge as superlative artifactual performance. There are three important components to this thesis. First, Allen proposes a novel theory of artifacts which avoids reductionism and instrumentalism and which, in turn, introduces a pre-linguistic space of artifactual performance which is already normatively and conceptually articulated. Second, he uses this intervention to undermine the priority of propositional knowledge to non-propositional knowledge, since according to his account, propositional knowledge arises only after the pre-linguistic artifactual realm has been established. Third, he resituates the history of knowledge as a history of everchanging artifacts, inclusive of our material technologies, linguistic concepts, and mental states.

Regarding his position on artifacts, Allen is most influenced by Henri Bergson and his defense of *homo faber* over *homo sapiens*.[25] As a notorious opponent of rationalism, Bergson ([1911] 1998) argues that human intelligence is not tied to a faculty of reasoning. Rather, *intelligence* is "the faculty of manufacturing artificial objects, especially tools to make tools, and of indefinitely varying the manufacture" which he then contrasts with the instinctual drives of animals (139).[26] Allen similarly defends a conception of *homo faber* which treats the manufacture of artifacts as a distinctly human achievement. Like Bergson, Allen sees tools or artifacts as already

[25] In a later article, Allen (2013) makes this influence more explicit by comparing Bergson's account of knowledge with that of the classical pragmatists. He also extensively analyzes Bergson's philosophy of technology in *Living in Time: The Philosophy of Henri Bergson* (Allen 2023, Ch. 4).

[26] This distinction of tool-use from instinctual behavior is a view Bergson and Dewey share. See Dewey on impulse (HNC, MW 14: 65–66). The difference is that Bergson believes that intelligence must be supplemented by *intuition* or "instinct that has become disinterested, self-conscious, capable of reflecting upon its object and of enlarging it indefinitely" ([1911] 1998, 176).

embedded in a system of relations with other artifacts. In his terminology, *tools* are "designed and made, usually by others, to facilitate action" (Allen 2008, 2). This separation of intended design from individual use means that tools are meaningful objects outside of and prior to any individual intended use. When we use tools, they are "already artifacts when we pick them up, already products of others' artifice with other tools, reciprocally linked to other tools and other people in an economy of actions, agents, and artifacts" (2008, 3). This "intersubjective intentionality" separates human tool use from other kinds of animal behavior (2008, 3). When humans use a tool, it is both part of a "technical economy of other tools" and a "social economy of complementary action" (2008, 72).[27]

These technical and social economies are lacking in other forms of animal behavior. Bees and beavers—in building their hives and dams—develop *quasi-tools* which are a result of in-born behaviors rather than intersubjective intentionality (2008, 72). Animals such as chimpanzees utilize objects in a manner much closer to human tool use, but Allen thinks that these behaviors still fall short of being full-blown tool use. When a chimpanzee uses a stick to "fish" for termites, it uses a *proto-tool*; the object is utilized as a result of learned behavior rather than inflexible instinct, but this behavior lacks the relational dimensions of the technical and social economies (2008, 69). Unlike the hammer, which remains a hammer even when no one uses it, the chimpanzee's stick loses its status as a "tool" the moment it is discarded.[28]

This taxonomy leads Allen to develop a non-rationalist and non-intellectualist account of normativity which centers around human artifice. To use an *artifact* is to engage in a *performance,* where performance demarcates "normative accomplishment rather than an instinctive, species-specific behavior" (2008, 49). When humans use tools, they are using objects which have been designed by others, and which have been or will be used by others. These tools are articulated against a background of other tools. For example, hammers, nails, and wooden boards maintain their individual meaning in relation to each other as parts of a particular performance and as parts of a history of artifacts, a history which also includes those objects affected and which in turn

[27] There is a Deweyan thread in Allen's views, which Rorty notes in his response to Allen. Rorty tells Allen that "Dewey would have liked the idea of replacing epistemology with a general theory of artifacts" (Rorty 2000a, 237).

[28] Ana Cuevas-Badallo (2021) has recently provided a different pragmatist taxonomy of technical capabilities in non-humans. She follows Tomasello (1999) in attributing to human technical practices the unique aspect of the *ratchet effect*, a phenomenon where a singular kind of artifact is subsequently modified in its design by different users across generations.

affect, our performances with tools (49).[29] Because of this system of interacting human actors and artifacts, tools develop a "functionless functionality," where they become re-embedded within new problems and performances which undermine any strict essentialist notion of their individual function (75). This protects Allen from charges of instrumentalism, since he argues that "a tool can't be an instrument without being more than an instrument" (17), that no tool is constrained to a particular function across uses. In this manner, Allen's notion of tools is more than an account of desire-satisfaction by use of objects external to the body. Each tool is crafted for a purpose but then rendered plastic by a plurality of possible performances.

In understanding human tool use as already situated with a collection of evolving practices, Allen envisions a human history where a field of material objects was already available for meaningful performance prior to humanity's possession of linguistic or discursive intentionality. This field of objects is the key to Allen's attack on the biases of epistemology since it undercuts the primacy of discursive intentionality and revises socio-practical intentionality to avoid charges of reductionism and instrumentalism. In Allen's view, knowledge is "superlative artifactual performance" (2004, 67). This describes any case where an individual produces something novel with given artifacts. The qualifier 'superlative' is highly revisionist as it disqualifies most habitual know-how as instances of knowledge. Knowledge is neither know-how nor knowledge-that; it is an a*chievement* which comes from exemplary or novel demonstrations. As Allen puts it, "it is not acquired merely by opening your eyes, or by doing something anybody can do" (68). It is more than a cognitive process or mental state.[30]

Just because knowledge is not mere know-how, does not mean that it is always knowledge-that. Allen thinks that the very division of knowledge into knowing-how and knowing-that "confuses media and expressions of knowledge with ranks and species of knowledge" (2004, 67). Sometimes knowledge takes a propositional form and sometimes it does not. Despite rejecting the distinction, Allen nonetheless treats propositional knowledge as secondary to other kinds of knowledge, at least in the historical sense that language emerged only after the use of tools. For

[29] Allen has an idiosyncratic view of artifacts which would count such things as pollution, the ozone layer, and the human brain as artifacts given that they are the effects of human performance. Perhaps, to avoid misuse of the word, we should think of these phenomena as "artifacts" as either items of the natural world with normative influence or as unintentional byproducts of utilizing tools.

[30] I take the 'superlative' part of Allen's position on knowledge to be the weakest part of his view since he openly states that he cannot provide any criteria for how to determine when a performance is superlative (2004, 69).

Allen (2004), propositions inevitably become "artifacts, things we have to know how to use before anything we do with them can be knowledge" (17). Since the introduction of true propositions amounts to piling up linguistic artifacts on top of the non-linguistic ones, there is no sense that any set of true propositions can "get at" the content of the initial artifactual knowledge nor can they cut beneath the artifice to get at unconstructed reality. Using the Brooklyn Bridge as an example, Allen explains:

> When knowledge begins in artifacts there is no guarantee that it can be translated into true propositions. No set of propositions, however true, is equivalent to the knowledge the Brooklyn Bridge embodies. I'm not saying the knowledge is ineffable. An engineer can tell a lot about the bridge, how it works, why it is significant engineering. But retrospective analysis is in no way equivalent to the knowledge the bridge expresses, which is as singular and nonrepresentational as the bridge itself is. (2008, 37)

This view is that artifactual knowledge cannot be fully captured by linguistic description or explanation. As Allen elsewhere states, products of human intentionality are "overdetermined" (2008, 5). Technology, works of art, and other artifacts "do not speak—they look or sound" (3). They affect us "without discursive, linguistically mediated reasoning about what they are, or mean, or might be good for" (3).

This rejection of the primacy of propositions does not lead Allen to fully dismiss the importance of discursive reasoning. Language comes into the picture only after humans have already developed systems of tools. It serves as a robust instance of a *cognitive artifact,* an artifact which makes use of external environmental resources to extend thought (2008, 25). Writing instruments such as papers and pens, representational media like photographs and MRIs, and time-keeping devices like clocks and schedules equally serve to distribute and embody our thinking into items external to our bodies (25–27).[31] Language, however, is different from these other tools in that the presence of syntax creates an open and flexible system which allows us to move beyond the immediacy of experience to describe negations and possibilities.

Drawing together the different parts of his program, Allen's re-conception of language as a kind of artifactual knowledge radically transforms our understanding of the history of knowledge. Against Rorty and Brandom's idealist picture of history, where knowledge is an accumulation of new beliefs, Allen recasts the history of knowledge in material terms as a

[31] One of the most famous assertions that our thinking extends beyond our body and into our artifacts is given in Andy Clark and David Chalmers' thesis on the extended mind (see Clark and Chalmers, 1998).

generation and accumulation of artifacts. In this new picture, even the manner in which our internal mental states change is likened to the change in the use of particular artifacts—linguistic and non-linguistic—within our practices. Allen states:

> The history of knowledge is not a story of change in what we do internally—a change in methodology, epistemology, or logic. What changes are the things we think and know with. What's really different about knowledge today, or knowledge in modern technoscientific societies, is that there's an awful lot more artifacts to begin with. (2008, 39)

Allen's picture moderates the typical Whiggish conception of progress with a skeptical concern about novel effects. Knowledge in modern societies is more advanced than in the past, not because we now have more true beliefs, but because we have accumulated more artifacts and thus have more resources to *think with.* But at the same time, *more* does not equate to *better.* The presence of new tools also yields new unintended consequences, some of which have the potential to detrimentally impact the globe. One benefit to Allen's extension of knowledge to include artifactual performance is that he can better address these kinds of problems which are simply outside of the scope of Rorty's pragmatism. However, though Allen mounts a compelling attack against both Rorty and Brandom, there are some holes in his position which render him vulnerable to linguistic backlash.

### 2.4 – From Neo-Pragmatism to Artifactual Pragmatism

Faced with his two critics, Rorty resituates his own position against both Brandom's rationalist extension of the linguistic turn and Allen's revisionist theory of knowledge as artifactual performance. Holding fast to his longstanding views, Rorty agrees to Allen's framing of knowledge as artifactual in order to undermine Brandom's representationalism while also agreeing with Brandom's linguistic normativity. This set of views largely tracks Rorty's earlier positions, but by reiterating a more robust theory of artifacts, Rorty's own endorsement of linguistic priority can be questioned.

First, in response to Brandom, Rorty (2000b) agrees with the claim that there are some instances where an observer can attribute knowledge or justification to an individual without that individual being aware of their knowledge.[32] But he reiterates that this is in accord with his anti-

[32] This is already a quite radical admission by Rorty. He too could say Galileo was justified in believing that 'The Earth revolves around the Sun' even though contemporaneous social practices would not have justified this belief. But Rorty's variation denies that it is *the facts* that justified Galileo's belief. He would prefer to say it is *us*, in our present standpoint, who attributes justification to him.

representationalism. Deflating Brandom's position on facts by means of analogy, Rorty states, "Brandom's strategy of treating truths, and facts, as true claimables, seems to me like saying that the rules of baseball were there, but unexpressed, before baseball was played" (184). Rorty grants that even if one accepts this position—that there existed true claimables before the existence of claims just as there existed possible baseballs before baseball—there is little practical difference in the matter, for claimables, as *possible* objects of claims, are not *actual* objects capable of standing in truth-making relations to claims. The result is akin to Molière's infamous assertion that opium puts people to sleep because of its dormitive power (185); there is no more explanatory success in Brandom's position on claimables than in Rorty's own commitment to a deflationary position that there are no extra-linguistic truth-makers.

One way of putting this response is that Rorty is suggesting that Brandom's appeal to 'claimables' is merely an adoption of a vocabulary which will permit relations between claims and facts. In this vocabulary, one takes up 'claim' as a truth-evaluable sentence and one posits 'claimables' as those states of affairs that "make claims true independent of any actual claims." But since, both the claims and the claimables are only understood in relation to each other as posits of the adopted vocabulary, there are still no real word-world relations—no relations between intra-vocabulary and extra-vocabulary objects—in which a robust notion of representation can be understood. As a result, Rorty sees Brandom's position as collapsing back into an instrumentalism where "coping" replaces "copying" (2000b, 185); Brandom has merely taken up a *set of tools* which allows him to agree with representationalism at a surface level.[33]

Against Allen's criticism, Rorty's concession is more radical in that he readily admits to holding a propositional bias which is suffused throughout his entire oeuvre. In one of the most surprising passages of *Rorty and His Critics*, Rorty revises his earlier positions on language use, extending assessments of reliability to all parts of our practical know-how in a manner that draws more fruitfully upon Dewey's account of the means-end continuum and Heidegger's position on tool-use. He writes:

---

[33] Rorty (2000c) does concede to Björn Ramberg that there is a sense in which our linguistic assertions "get the world right," but he does not like appealing to sentence-shaped entities called facts to do the explanatory work of this relationship to the world. He instead likes saying that we get the world right as a result of a messy collaboration with the physical world and the social order rather than saying correctness is a result of the one-to-one relationship between claimings and facts. See Yvonne Huetter-Almerigi's (2020) "Two Forms of Realism" for a masterful exposition of why Rorty concedes to Ramberg but not to Brandom.

> Dewey, like Heidegger, spent a lot of time saying that it pays, for certain purposes (for instance, getting rid of epistemology) to view the Vorhanden as a species of the Zuhanden. Perhaps I should have spent more time doing the same. If I had, I would have said that sentences, skills, and disciplines (such as historiography) can all be treated as artifacts. All three can, like the spear and the dam, be evaluated in terms of reliability. But I would also have said that, so treated, artifacts are continuous with organs: the beaver's teeth and its dam, the human hand and the tools it grasps, can all be so evaluated. (Rorty 2000a, 238)

This modification concedes to part of Allen's project. Rorty agrees that, for the purposes of naturalization, he should have extended the metaphor of the tool to all human domains. As hinted in the passage above, this concession brings Rorty much closer to Dewey and the earlier pragmatists, in particular Dewey's treatment of artifacts and our bodies as tools which are assessable in relation to identified ends (HNC, MW 14: 22).

But Rorty's consideration of artifacts does not imply his endorsement of Allen's alternative definition of knowledge as superlative artifactual performance. Rorty finds Allen's program (rightfully) too revisionist. He thinks we should give up the 'superlative' qualifier and maintain the distinction between knowing-how and knowing-that by appealing to the unique function of linguistic artifacts relative to non-linguistic ones:

> On one hand, it is hard to improve your non-linguistic artifacts unless your peers can deploy critical remarks about their design; for lack of linguistic artifacts to use in making such remarks, the beaver dams of our own day are no better than those of old. On the other hand, it is hard to have the leisure for language-building if you lack non-linguistic artifacts with which to defend yourself against the climate and the predators. One can see why the two kinds of artifacts are likely to have been produced around the same time, and to have developed in tandem. (Rorty 2000a, 239)

Rorty's resolution of the tension carefully modifies his own views without departing from them. In particular, he uses Allen's artifactual framing to reiterate his anti-representationalist theory of language and Brandom's rationalism to preserve an account of normativity that attributes responsibility only to language users. Yet in his conservatism, Rorty inadequately addresses some of the important theses that Allen and Brandom have defended.

First, Allen's (2004) position suggests that artifactual knowledge is distinct from other forms of animal behavior in that artifacts are component parts of an extended economy of tools. Rorty's response fundamentally ignores Allen's main point by collapsing the distinct normativity of non-linguistic artifactual performance into a behavioralist stimulus-response account of animal action. However, according to Allen's taxonomy, the beaver's dam is strictly-speaking an

instinctual behavior, rather than an intentionally manufactured artifact. Despite conceding to Allen's framing, Rorty does not consider that Allen's *homo faber* is more than just another reliably responsive animal.

Second, Rorty maintains his deeply instrumentalist account of language use in opposition to Brandom's and Allen's anti-instrumentalism. Allen maintains his anti-instrumentalism by appealing to the technical and social economies, by conceiving of tools as lacking essential functions, and by adding the 'superlative' qualifier to knowledge. Without these resources, Rorty risks conflating knowledge with organic adaptation to an environment. Following Brandom, Rorty treats our ability to wield sentences as the deciding factor between knowledge and adaptation, but if he conceives of language as just a meta-tool for modification of other practices, then there is nothing normatively autonomous about the space of reasons in relation to other practices. From Brandom's perspective, the worry is that Rorty lacks any real account of normativity, reducing the norms of our reasoning to some non-rational activity. This risks, for example, conflating the *truth* of one's statements with how successful these statements are as instruments for coping with our environment. As with Allen's *homo faber*, Rorty risks treating Brandom's *homo sapiens* in reductionistic and behavioristic terms.

Finally, Rorty's compromise maintains the relationship between knowing-how and knowing-that on the basis of linguistic versus non-linguistic artifacts, but his explicit adoption of a broader class of artifacts destabilizes the neat distinction of linguistic justification from non-linguistic coping. Part of Allen's position on artifactual knowledge is to show that certain artifacts such as artworks and photographs do not strictly function as posits of knowledge-how nor do linguistic artifacts strictly function like knowledge-that. Although Allen ultimately chooses to discard the distinction, he highlights how the difference between knowing-how and knowing-that is often confused with the distinction between various media or vehicles which express knowledge. Perhaps both distinctions are valid—there is a sense in which non-linguistic artifacts function *rationally* like knowledge-that and in which linguistic artifacts function as knowledge-how. In these cases, a variety of media can play both roles. Indeed, it seems at least plausible that sometimes *homo faber* uses linguistic artifacts and sometimes *homo sapiens* uses non-linguistic ones.

## 2.5 – An Outline of Artifactual Pragmatism

I argue that the solution to the dilemma is to accept that *homo faber* and *homo sapiens* each express some important feature of normativity, each attributable to the different functions of our artifacts. Rorty comes close to this position in his later developments but ultimately fails to grasp the important features of Brandom's and Allen's views which would have helped him to revive an adequate pragmatist theory of normativity and objectivity. My own position, what I call *artifactual pragmatism,* starts with the view that pragmatism is fundamentally a theory about the meaning of *artifacts*. It then derives, from this view, positions on *artifactual normativity* and *artifactual reasons*.

As a brief definition, artifacts are manufactured objects which derive their meaning from intersubjective use by social actors. They are normative phenomena in the minimal sense that they are defined by their socially prescribed functions, even though they are not necessarily bound to such functions—for instance, a wrench can be used *as a hammer,* but this does not mean that the wrench becomes a hammer. Artifacts are also conceptually structured phenomena. They are placed into classes by means of their function, and these functions are subsequently linked in compatibility and incompatibility relations with other artifacts. For example, a Phillips head screwdriver is functionally compatible with a Philips-head screw but not with a flathead screw; meanwhile, a flathead screwdriver of the right size can be compatible with a Phillips head screw though with less grip than the Phillips head screwdriver. These functional instances of compatibility and incompatibility structure the "technical economy" in a manner that already sorts artifacts into classes even though these classes may not possess a common name. Nonetheless, when these artifacts are incorporated into social performances, other individuals may reward or punish those usual or unusual performances which satisfy common ends.[34]

This view fundamentally opposes Rorty and Brandom's linguistic conceptions of normativity. According to artifactual pragmatism, one can be a concept-mongerer who can attribute moral responsibility to others and be equally attributed responsibility by others without

[34] There is a host of non-representational theories of concepts which could explain the presence of artifactual classes without linguistic classification. In the vein of pragmatism, Johnson and Tucker (2021) see concepts as arising from the affordance patterns provided by perceivable objects. I think this position maps closely to my own, with the addendum that I think concepts can only function *normatively* if they are expressed in a social practice. Johnson and Tucker's examples of hammers and chairs work well to show that concepts require that an individual experience an object's affordances in a manner which are aligned and coordinated with the expected affordances of others.

being able to provide reasons. In fact, it seems like the moral stance we adopt towards young children and pets accepts that these beings have correctly mastered the rules of some of our social practices in a manner where intentional wrongdoings can be distinguished from unintentional errors on the grounds of non-linguistic performances. What I concede to Rorty and Brandom is that being a sentient creature alone may not be enough to attribute possession of concepts in the normative sense, since there are issues related to positing a "private language" inside of a creature's head.

When creatures are integrated into shared practices, they develop behavior which adopts the standards of the general community. This matters for locating degrees of normative agency among individuals. We may rightfully scold or sanction a well-trained dog who bites a child while also admitting that such sanctioning would be irrational or even pointless in relation to the actions of a wild wolf who did not receive the proper training. The difference is that the dog's training has led it to demonstrate possession of the ability to discriminate between a class of biteable artifacts such as food, balls, frisbees, and chew toys from non-biteables like humans and couch cushions.[35] Given the vast plurality of relationships of combability and incompatibility between these objects, the dog seems to have mastered rules relating to the correctness of its actions even though she cannot give reasons for following these rules.[36]

One way of putting my position is to say that social creatures do not need to behave in a manner that produces explicitly propositional content to have norms. Children, animals, and those without physical or cognitive capacity to provide reasons, are still bound up in systems of rules which they may grasp. More importantly for my project, those who do participate in the space of reasons possess linguistic norms which are constantly entangled with other object-related norms. In this manner, the norms of language—as the use of words or sentences—are inseparable from

[35] Notably, this view is not what Brandom (1994) terms an "erotic theory" of classification, where an organism's drives and desires carve up the world into concepts (86). Artifactual pragmatism holds that *social practices* give artifacts their normative and conceptual status, not the individual psychology and activity of each organism.

[36] Here, Dewey's account of *training* versus *education* is helpful, since he distinguishes between a creature acting from the subjective standpoint of desire-satisfaction versus the intersubjective standpoint of some common practical end (DE, MW 9: 17–18). One potential problem with this account that treating a norm as merely acting with respect to a common end might confuse trivial normative violations (e.g. using a fork as a backscratcher) with substantial ethical wrongdoing (e.g. stabbing someone with said fork). However, I think the difference between the two kinds of violation is explained by different kinds of sanctioning. A person misusing a fork in the former sense might receive disgusted looks from onlookers while a person misusing the fork in the latter sense might be forcibly restrained or arrested.

norms relating to non-linguistic conceptually structured items. This is to say that artifactual pragmatism agrees with Sellars in treating habits as social rules and with the Quine-Davidson suspicion towards treating language as a self-contained axiomatic system. The linguistic rules one knows cannot be disentangled from the practical rules one *embodies* in knowing one's way around a world populated by artifacts.

By extending normativity to non-rational beings and practices, I do not intend to undervalue the status of the space of reasons. The dog cannot easily modify the rules of their practices without breaking them and being subject to sanctioning by others. Non-rational modifications of practice can occur in terms of counter-conduct rather than through rational persuasion or social deliberation, but what the space of reasons allows for is the practical ability to provide reasons for or against the continuation of a particular practice and the theoretical ability to talk about parts of the world which do not have any bearing on the success of a practice. In this manner, Rorty is correct that the space of reasons practically involves modification of non-rational practices in the sense of using linguistic artifacts to perfect non-linguistic ones. And Brandom is correct that the space of reasons theoretically provides us with true sentences (and true beliefs) whose truth status is not directly determined by the practical ends of any organism or any other kind of social practice. The mistake shared by both is to make linguistic capacities a prerequisite to genuinely normative practice, that is, to deny the inherent normativity in the use of artifacts.

### *2.5.1 – Artifactual Normativity and The Space of Practice*

One of the core insights of artifactual pragmatism is to carve up the contemporary landscape of pragmatism differently from other adherents of Sellars. For this purpose, Daniel Dennett (1987) provides a helpful trichotomy. Dennett argues that, for the purposes of predicting the behavior of some system, we can take up three possible stances: the *physical stance* which treats the system as some set of basic physical components acting according to laws, the *design stance*, which treats the system as acting toward some prescribed function, and the *intentional stance*, which treats the system as possessing intentionality in the form of beliefs and desires (1989, 15–18).[37] The idea of the three stances imperfectly aligns with my notion that there is some dimension of human practice

[37] Following Dennett, some philosophers specifically see artifacts as objects of design which are neither strictly causal nor rational (Floridi 2011; Kroes 2012). I hesitate to speak of "design" due to general concerns about importing the language of biological function to also describe the social. Dennett, for instance, is comfortable attributing the design stance to human organs in a manner that naturalizes function (1987, 17).

which is neither captured by causality nor by attributions of belief. What Rorty and Brandom miss, relative to the three stances, is that we live in a world populated by artificial objects which are ascribed functions and norms by our practices. The sharp distinction between the *space of causes* and the *space of reasons* cannot be maintained in light of this class of objects and practices. The pragmatist should therefore define a new space which is neither purely causal (that is, non-normative) nor purely rational in scope—a *space of practice*.

This middle space is inspired by similar metaphors across the history of philosophy. In Aristotle's *Metaphysics*, it is the space occupied by the craftsman, who is situated above the experiencing animal in his grasp of the forms but below the philosopher who uses the forms to investigate the causes of things (981b25—982a5). And in the work of Hannah Arendt, the middle space of human activity is designated as *work* and is situated between *labor*, designating our biological processes, and *action*, designating the political life constituted by words and deeds (1958, 7). In both of these conceptions, the middle space is occupied by *homo faber* as the fabricator, the designer, or the maker of artifacts. In both accounts, the highest form of life is dependent upon the stability and durability first established by human artifice.

Within pragmatism, the basic impetus of the space of practice has its origins in several positions stemming back to the work of James and Dewey and extending into the work of the later-Wittgenstein. My terminology of the space of practice borrows from Simon Blackburn (2019), who has proposed, as a counter to Sellarsian rationalist pragmatisms, that there exists a "space of intelligent practice" which "smoothly amalgamates doings and reasonings" (119). This proposal exemplifies what Brandom takes to be a general Wittgensteinian strategy toward normativity: discursive intentionality is treated as a subspecies of a broader practical intentionality and practical intentionality is, in turn, redescribed in terms of social practices (Brandom 2019, 88). Where Blackburn's Wittgensteinian picture departs from Brandom's rationalism is that he argues that Wittgenstein's various language games are genuinely normative since the builders are "intelligently engaged" beings caught up within practices (115).[38] Though Blackburn holds, in-

[38] I suspect Brandom's Kantian inclinations would lead him to dismiss this use of 'intelligent'. However, other Kantians, such as Christine Korsgaard (2018) are comfortable with distinguishing rationality from intelligence. In Korgaard's terms, an *intelligent animal* is "characterized by his ability to learn from his experiences, and to solve problems by taking thought. He is able to extend his repertoire of practically significant representations beyond those with which his instincts originally supplied him" (39). In contrast, a *rational animal* "is (at least sometimes) aware of the grounds of her beliefs and actions. [...] This consciousness puts her in a position to evaluate those grounds and decide whether or not they count as *good reasons* to act and believe in the ways that they tempt her to act and believe" (39).

line with Brandom, that the builders cannot make the norms of their practices explicit, he does not think this situation bars them from being able to follow self-standing norms. He writes:

> The builders are not shown deploying norms and normative language. But there is, of course, space for characteristic normative behaviour to enter their primitive world —after getting beams by saying "beam" builder A is likely to be annoyed if B brings what A regards as the *wrong* item such as a slab instead of a beam, just as a teacher introducing numerals may feel like *giving up* on the learner who cannot get the sequence right. But this is going to be true whenever someone disappoints expectations or fails to do his part in the joint exercise of a technique. (Blackburn 2019, 115)

It is important that Blackburn compares the builders' capacity to follow rules as an equivalent case to Wittgenstein's instance of rule-following with respect to learning arithmetic. The problem posed is why there is a need to treat linguistic behavior (or in this case mathematical behavior) as separate from the domain of non-linguistic behavior. As Blackburn puts it, the central challenge to Brandom's conception of practice becomes "whether it is wise to dissociate the essential nature of linguistic practice from the essential nature of the object–related practices and activities" (117). Removing this distinction shows that the builders already operate according to rules in world-involving ways.

If Blackburn is right, then the classes of practical objects involved in the builders' performances—slabs, color samples, vocal calls—function as objects of social performances which can be evaluated in terms of correctness. Object-related practices themselves sort objects into public, normative classes depending upon their instrumental value to that practice, and for a being which participates in a variety of practices, a range of artifacts from across these practices are related to each other in a manner that produces identifiable uses and norms for each artifact. Brandom's point still holds that the builder who produces the wrong slab makes a mistake according to a subjective norm held by the group, but his presentation of this kind of normativity risks obfuscating their immanent existence in a world which pushes back. The slabs and samples themselves possess mind-independent properties which have been indexed and put to use within the practice. Though the builders cannot state true sentences, their non-linguistic practices already involve feedback. This is because their practices implicitly, rather than explicitly, index objective properties which then play a causal role in the success of the builders' performances. What Blackburn shows is that the pre-Sellarsian pragmatists are united, not in their narrow instrumentalism or their bald naturalism, but rather by their creative appeal to tools and instruments

as useful metaphors for explaining how our basic bodily practices are already intermixed with the natural environment.[39]

The notion of a space of practice closely follows other developments within pragmatist-inspired positions on science and technology. In the philosophy of technology, figures such as Langdon Winner (1986) have long drawn from language games as an explanatory device for how our use of tools and artifacts are "forms of life." More recently, Mark Coeckelbergh (2017) has gone further in this arena by introducing the concept of the *technology game* to explain how the meaning-in-use dimensions of Wittgenstein's philosophy of language can be readily extended to our use of artifacts. Coeckelbergh admits that one reading of Wittgenstein as an anti-representationalist "downplays the material object," but he also believes that "we can bring back the object and technology by turning around his metaphor of the toolbox" (29). Doing so not only escapes linguistic idealism, but it correctly "highlights know-how, practice and the lifeworld" as core dimensions of Wittgenstein's epistemological outlook (29).

Various pragmatisms have also been proposed which center a broader conception *social practice* as the basis for a pragmatist normativity. In defense of the classical pragmatists against Brandom, Steven Levine explains that in their account "practices—i.e., habits and bodily skills—are neither fully discursive nor merely causal" (2012, 128). Colin Koopman's transitionalist pragmatism equally challenges Rorty and Brandom's narrowly linguistic construal of normativity by reconceiving epistemology to be an epistemic relation between our anticipated projected aims and the realization of these aims through practice (2009, 109).[40] And within the philosophy of science, Hasok Chang (2022) develops a pragmatist position which first identifies a plurality of game-like *epistemic activities* which combine into institutionalized subfields or *systems of practice* (15). Crucially, these three proposals share Allen's skepticism towards propositional knowledge. Levine thinks habituation captures a set of pre-reflective yet normatively-articulated embodied

[39] It is beyond the scope of this chapter to provide an exegetical defense of the earlier pragmatists. However, I provide part of this defense elsewhere in my article "Technology and Social Practice: Reassessing the Language-as-Tool Metaphors in Dewey and Wittgenstein" (Caplan 2025). There I argue that Dewey's conception of habit and Wittgenstein's signature language games are appropriately analyzed according to Blackburn's "space of intelligent practice" in a manner that warrants re-attention by both linguistic and experiential pragmatists.

[40] Koopman (2014) later formulates his project as conduct pragmatism, which locates both language and experience within the broader field of conduct. This position amounts to locating normativity within practices that lie beyond the linguisticism of Rorty and Brandom while also downplaying the pre-practical qualitative immediacy present within Dewey.

skills (2012, 138). Koopman argues the same position with attention to the objects of our practices. In his view, epistemic practice "often requires the skillful use of a variety of tools at once (sentences, hammers, meals, and strategies for social cooperation)" (2009, 128). Chang (2022) too readily admits that "activities are not reducible to propositions in any straightforward sense" (16).[41] These three positions assert, along with Winner and Coeckelbergh, that the pragmatist cannot start from Rorty and Brandom's linguistic standpoint.

Artifactual pragmatism captures the core theses of these various proposals by establishing a space of practice to which activities, practices, technology games, and the like can be mapped, but it distinguishes itself from these other positions in that it argues that the space of reasons emerges from the intersection of these already-normative practical activities. This is in contrast with social pragmatist accounts that too readily associate the game of giving and asking for reasons with all intelligent practice and with some Sellarsian commentators who argue that an individualistic notion of *embodied practice* underlies or overrides the space of reasons.[42]

My notion of the space of practice is thoroughly Wittgensteinian in that it is both *embodied* and *social*. It comes closest to the later Davidson, who argues that a being can only think, in the specifically normative sense, if it "physically resembles a person in important ways, and has an appropriate history" (1990c, 86). Davidson contrasts his position with those which attribute understanding to an artificial system solely by its linguistic competency, for instance those of Alan Turing (1950). Davidson expresses that the "physical realization of a program" matters to its capacity to understand the world (1990c, 86). Yet he is not defending the view that this physical realization is necessary for *sentience*; he is not even concerned with that question.[43] Rather, he

[41] Through, in contrast to my reading of Sellars in the Introduction, Chang sees Sellars' social-functional treatment of persons and their doings as providing an avenue out what he calls "the flatland of propositions and their logical relations" (2022, 32).

[42] Susan Hurley (2003) suggests the "space of intentional actions" should replace the space of reasons in order to account for the intentionality present in animal motor action. Carl B Sachs (2013) suggests a "bifurcated intentionality" where the discursive intentionality of Brandom is paired with a "somatic intentionality" inspired by Merleau-Ponty. Similarly, Danilo Manca (2022) defends a phenomenological and enactivist position in which the living body constructs the space of reasons out of its affordance-rich environment.

[43] John Searle (1980) famously endorses the position that only a biological entity can possess meaning. The general Wittgensteinian response, which Davidson would share, is to suggest he has conflated sapience and sentience, of knowing the meaning of a language versus being aware and capable of feeling pain and pleasure (see Gubelmann 2023; Søgaard 2023).

provides a condition for normative agency, one in which our social engagements in the world provide a set of common objects which precede our mastery of a language.

What artifactual pragmatism adds to the Davidsonian picture is a conception of engagement with the world that is already structured in normative ways. For all of history, an entity's possession of a language was inseparable, not only from its mere existence in a physical world, but from its being an actor within a plurality of practices or forms of life.

The space of practice describes our intelligent conduct, our learned behaviors, our capacity to re-hash and re-mix antecedent ends, but it should not be confused with our ability to feel pleasure and pain. It instead captures what it would be like to participate in a plurality of different self-contained practices without positing beliefs, engaging in scorekeeping, or redirecting other's activities by means of rational persuasion. It is the space inhabited by children and clever animals (and possibly also embodied artificial entities), those beings who know their way around the world by knowing their way in and around social practices, beings which know how, when, and where to leverage particular learned abilities and skills. It is also the space inhabited by rational creatures when they participate outside of reflective activity—the gymnast or concert pianist whose performances are dependent upon "muscle memory" and are consequently impeded by the "noise" of the mental. These persons still follow norms in their performances, even if the expression of these norms can only be made public in the activity of the performance itself. In the space of practice, we are conceptual bricoleurs, creatures of habit, or clever craftspersons who shift around the given materials of the world into new artifactual arrangements in order to produce new kinds of ends.

The space of reasons is one special kind of practice among many. We enter into the space of reasons, as a kind of meta-game, whenever we reflect upon these other practices, when we track the beliefs of others, and when we ask ourselves or others *for justification*. The space of reasons cannot function independently from the field of practices—this would be like trying to give true claims without being able to speak, write, or sign. This space cannot function independently from the materials of practical activity and often functions as a rational game overlaid over practical activity, as when a poker player follows the rules of the game while also discerning the intentional states of his opponent. Other times, the space of reasons spontaneously emerges and dissipates as practical activities are problematized and resolved communicatively by their participants, as when

the pianist stops her train of thought to reconsider her fingering in her head or when a soccer player contests the decision of a referee.

The space of reasons describes our existence as conceptual engineers, as inquiring minds, or as careful conversationalists who use the materials of the world to craft justifications and produce new kinds of persons. Overtime, all human practices have been modified by this meta-practice. But though the space of reasons may affect all practice, its norms and functions remain distinct. [44] The rules of language, logic, and argument are as unique to their corresponding practices as the norms of football or baseball, even if a total history of these practices would see them as interacting parts of a whole system of norms whose distinctions have been set pragmatically.[45]

### 2.6 – A Puzzle About Reasons: Signs, Artifacts, and the Matching Game

In setting up three distinct spaces, the puzzle of external friction is shifted. Regarding the relation of practice and cause, I have offered a tentative answer, in the introduction, that social practices index properties of things in order to transform them into some artifactual kind. The relationship does not seem too mysterious—outside of social practices, tables and chairs are identical to their similarly-shaped physical constituents. It is when creatures intersubjectively use an object according to a common end they turn the physical constituents into an artifactual kind,

But how do artifacts relate to reasons? Above, I described how a practical game and a rational game may be performed at the same time. But there is a second kind of relationship where practical objects can function like sentences, claims, or premises.

As a conclusion to this chapter, I want to set up the basis for my argument in the next chapter. In specifying the distinction between the space of practice and the space of reasons, the distinguishing feature of artifactual pragmatism is a position on *artifactual reasons*. This is the

[44] The central argument, as to be elucidated in the next chapter, is not to deny that linguistic or inferential rules are required in order for the space of reasons to get going, but rather, that a variety of non-linguistic items can be integrated as conceptual content. For example, "The black circle is left of the black square" uses the conceptual vehicles 'black', 'circle', 'left' and 'square' in the same manner as '●■' uses the conceptual vehicles '●' and '■'. The vehicles 'circle' and '●' are different conceptual vehicles because they have different use conditions.

[45] For instance, safety rules in one sport may be used as a guide for the rules of another without radically merging the two self-contained practices (though there might be metaphysical arguments to the contrary). Neither would justification of the rules of a practice constitute a merging of the space of reasons with that practice. Where common sense dictates, the awkward distinctions of one practice from another are often set precariously; this is no reason for the Wittgensteinian (or in a more classical frame, the Aristotelian) to avoid the concept of 'practice'.

view that the non-linguistic objects indexed by our practical activities can function *as reasons* in the space of reasons even though they are not strictly speaking sentences, propositions, or states of affairs. The result is a view that better expresses Brandom's intended aim, that discursive practices be seen as "solid (even lumpy), in that they involve actual objects and states of affairs, as well as the deontic statuses in terms of which score is kept" (1994, 632).

The problem for Brandom is that it is hard for him to endorse this thick and lumpy world while dividing the world into claimable contents.[46] As I see it, Brandom's propositional bias leads him to adopt a particular vocabulary for entities that function as reasons which is distinct from that vocabulary he uses to describe the ordinary entities which causally interact with us.[47] However, collapsing these vocabularies into each other is not as easy as it might seem. If we agree with the classical pragmatists, Wittgenstein, and Blackburn that the capacities for learning a language are already bound up within object-related normative practices, we undercut Brandom's overly discursive worldview, but in doing so, we undermine the sharp *linguistic* boundary which Brandom has used to keep reason-giving distinct from other kinds of practical activities.

I defer to Dewey's distinction between *symbols* and *signs* as an avenue out of this dilemma, since it allows Dewey to endorse the linguistic boundary without converting the material or objects "at play" within the space of reasons into linguistic items. According to Dewey's terminology, *symbols* are "artificial signs" or words which themselves "provide no evidence of any existence" (LTI, LW 12: 57). They are the units that "make possible ordered discourse or reasoning." (58), and when organized into propositions, they are "reasoning in the sense of ratiocination or rational discourse" (111). Though Dewey might start to look like Rorty and Brandom in prioritizing propositional formulations as the material of reasoning, Dewey contrasts symbols and their meanings with *signs* and their *significations* relating to a class of "natural signs" (57). Such signs are existing objects or events which relate to other existing objects or events, as when dark clouds signify rain or when smoke signifies fire.

---

[46] D. M. Armstrong (1993) endorses the view that the same spatiotemporal part of the world can exist both as a mereological composite of smaller parts (as describable by physics) and a state of affairs capable of serving as a truth-maker for a sentence. The problem with Brandom's talk of state of affairs is that there is nothing preventing his system from devolving into a minimal ontology of claim-claimable relations.

[47] This discrepancy between an object-language and an evidence-language was already identified by J. L. Austin (1962) in the conclusion of his *Sense and Sensibilia*. Regarding G. J. Warnock's linguistic reading of Berkeley (perhaps a very literal form of "linguistic idealism"), Austin writes "the resulting question about how the evidence-language ('idea'-language) is related to material-object-language, which he [Warnock] tries to answer, is a question that *has* no answer, it's a quite unreal question. The main thing is not to get bamboozled into asking it at all" (142).

With this set of distinctions, Dewey admits that though language is required for what I am calling the game of giving and asking for reasons, the material of natural signs should not be confused with that of symbols or propositions. He writes that language "has conferred upon *existential things* their *signifying* or *evidential power* (LTI, LW 12: 62, my emphasis). In this manner, propositions are the tools for relating and utilizing non-propositional material *as evidence*. Propositional or sentential items do not exhaust the available material for reasoning, they are merely "a way of acting that effects discrimination and ordering of existential material in its function as evidential data" (263). It is linguistic symbols which allow non-linguistic material to *become significant*, to function as evidence, as when the relation of the terms 'smoke' and 'fire' transform the perceived thing of smoke into a *sign* of an unperceived existing fire (61–62).[48]

Dewey's insight is to reject both an overly internalist conception of justification as a system of beliefs and an inflationary metaphysics which appeals to special objects beyond those identified by practice. Against Rorty, Dewey connects the world beyond our skins in justificatory relations to beliefs. Against Brandom, Dewey denies treating the world as already organized into a propositional structure. Following Allen, propositions are instruments or artifacts which are added on top of an already conceivable world of ordinary objects.

One way of inserting Dewey's insight into the landscape of analytic pragmatism is to draw from Huw Price's (2011) reconstruction of the representational "matching game" as an *internal representation* between two parts of our social practices. Price admits to holding a position which looks like correspondence or a matching relation between sentences and the world, one where "each statement identifies its own truthmaker" (23). However, following Rorty, Price provides a deflationary and anti-representationalist version of the correspondence relation which asserts that, in the matching game, "the pieces themselves are given their shapes as structured artifacts in a natural world" (23). However, since Price thinks semantic content consists of "sentence-sized packets of factual information" (26), the reconstructed matching game is an instance of matching

[48] Tom Burke (1994) provides the definitive account of Dewey's logical theory, though I slightly depart from his treatment. Burke writes that "facts *are* propositions, of one sort anyway" (1994, 175). There is certainly ample evidence for this interpretation in Dewey's *Logic.* But I think Dewey distinguishes i) symbol-symbol relations holding between words from ii) sign-signified relations holding between existent objects of inquiry (LTI, LW 12: 60) in order to avoid asserting that the world is experienced as propositional content. I worry that any alternate interpretation leaves Dewey open to the revisionist linguistic interpretation provided by Rorty (1977). Alternatively, Steven Fesmire finds exegetical support to show that Dewey's notion of 'proposition' describes proposals for action or performative acts rather than truth-bearers (2014, 55–56).

sentences to posited sentence-sized chunks of the world, something like projected states of affairs rather than artifacts in the sense expressed above.

Artifactual pragmatism makes use of Price's matching game by materializing and pluralizing it. Price is correct to see truth-makers as "structured artifacts" just as Brandom is correct to see perceptual facts as "conceptually structured truth-makers" (2000c, 164), but I think it is mistaken to compare the conceptual structure of *things* to propositional contents, states of affairs, or claimables. In my alternative picture, artifactual normativity shows how there are a variety of common conceptual vehicles—hammers, wrenches, soccer balls, dolls—which are conceptual in their being *classified* by their *use* by social creatures. This alternative collection of non-propositional, non-sentential conceptual objects, provides a range of materials that can be introduced, wholesale, into practices of justification in just the same manner as words or sentences are introduced. Similarly, following Dewey, learning a language merely transforms perceived natural material into more fine-grained collections of signifying or evidential material; natural objects become, in some sense, "artifactual" by serving as *potential* objects of practice.

The result of this position, I will argue, is that the matching game can be reconstructed as one which permits a variety of non-sentential objects. In my alternative picture, sentences pass around sentence-sized packets of factual information in the same sense that smoke passes around smoke-sized packets of factual information. Therefore, to hold the belief that "If there is smoke, then there is fire" is to use a sentence-sized packet to relate a presumed fire-sized packet to a perceived smoke-sized packet. This practice is extended to map complex forms of social verification in Chapter 5; there, matching a true sentence to a photograph will be understood as an instance of linking a sentence-sized information packet to a larger photograph-sized packet. Likewise, matching a true photograph of a person to the directly-perceived subject is like matching a photograph-sized packet to a larger person-sized packet.

This view is unorthodox, but it follows a Deweyan insight that linguistic symbols are intermixed with an innumerable number of observed objects capable of serving as evidence without reducing the evidential material to language. And it captures the attitude of common sense, that practices of belief alteration often involve a variety of socially-sanctioned performances which are not reducible to the linguistic. In my opinion, we should confidently say such things as, "The video is justified by the corresponding photograph" without being held guilty of endorsing some mythic given or of confusing causality with rationality.

But this position must fight an uphill battle.[49] Treating the linguistic pragmatist as my interlocutor, what is required for my thesis to work, are three substantive parts that I address in the next chapter: First, I must explain how there are common objects O whose identification and classification are not dependent on language. For this, I draw from Donald Davidson's account of triangulation. Second, I must explain, if there are such objects, how and when can they serve as perceptual reasons without being more substantially sentence-like. For this, I draw from John McDowell's conceptualist account of perceptual content. Third, I must explain the difference between pointing out objects in a natural environment, the basis of Dewey's examples, from using specifically-designed instruments like thermometers or photographs. For this, I integrate a reliabilist theory of justification into the account. The result is a new pragmatist picture of justification which combines already-existing positions to assert that: 1) *triangulated ordinary objects* are the 2) *conceptual contents of perception* to which 3) *reliabilist theories of epistemic justification* can be readily applied.

[49] The position is roughly a deflationary variant of what J.J. Cunningham (2018) calls an "object truth-maker" view of perceptual reasons, one which holds "When I believe that the cube is red on the basis of my visual perception, my reason is the cube" (2018, 263), and which he attributes to Bill Brewer (2011). The deflationary mechanism in my version is that the perceived objects are still part of a network of justificatory relations. They do not substantively serve as truth-makers, only as non-doxastic or non-sentential justifiers—the cube justifies my belief that it is red, but unlike the state of affairs 'The cube is red', it does not make my belief true.

# CHAPTER 3

## REASONS RECONSIDERED

> We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way, by using the old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction.
>
> — Otto Neurath, "Anti-Spengler" (1921)

### 3.1 – Introduction

In an attempt to link a neo-pragmatist methodology to analyses of social networking sites (SNS), Dutch philosopher of technology, Yoni Van Den Eede (2010) has offered one of the most robust applications of Richard Rorty's philosophy to the ethics and politics of online communication. Against the strand of thought which finds Rorty to have been too flexible on truth and justification, Van Den Eede sees Rorty's ironism as a helpful disposition which prevents total submission to a mode of thought within algorithmically organized spaces. More importantly, he also develops a defensive strategy against accusations of Rorty's linguisticism, arguing that Rorty holds a broad conception of 'conversation' in which our quest for solidarity "cannot be restricted to semiotic or linguistic elements" (204). In particular, Van Den Eede writes:

> Not all communication on SNS is linguistic. But does that mean that posting a photograph and showing it to all of one's contacts is not part of a real conversation—in which these contacts, for example, can then 'answer' by uploading their own pictures, or links, or videos, or music, and so forth? It goes to show that the notion of 'conversation,' is and should be, very much in a Rortian way, interpreted broadly. Every act of self-expression, self-description, or every moving story (movie, picture, song) about human suffering takes place within it. (Van Den Eede 2010, 204)

In Van Den Eede's interpretation, Rorty's conception of conversation is not strictly an "epistemological' mechanism;" it also contains the limits of what can be utilized within the social sphere (197), and thereby encompasses all domains of practice, not merely the linguistic or the propositional.

This reading would protect Rorty from a host of challenges. As R.W. Sleeper (1985) puts it in epistemological terms, if Rorty cannot appeal to anything more than the web of beliefs and desires, then epistemology becomes a twisted and radicalized rendering of Neurath's ship—it is nothing but a shifting platform of rotting planks out on the open seas which must be constantly rearranged and replaced, and in Rorty's case, there is nothing to guide or direct us to safer waters. Faced with Rorty's radical pragmatism, Sleeper recounts a way out of our predicament:

> We are indeed afloat in Neurath's boat, but what we need at this point is not just some new planks to shore up the hull. We need maps and a compass. It was Dewey's achievement to have provided access to these needed aids to practical navigation. By rejecting them Rorty's pragmatism sets us needlessly adrift in perilous seas. (Sleeper 1985, 19)

Yet Sleeper sees the navigational aids provided by Dewey as the grounded components of his *theory of inquiry*: the situation and the generic traits of existence. In a different vein, I will argue that we approach Sleeper's above passage in more literal terms—Dewey's most important contribution to philosophical methodology is not his articulation of the underlying structure of the situation but his appeals to the diversity of tools and instruments, such as the map and compass, which can be incorporated into our epistemic practices. As such, instead of appealing to a generalizable structure which underlies the web of beliefs and desires, I suggest that we expand our overall conception of the web itself to incorporate nonlinguistic objects, treating the constituent parts of our initially-nondiscursive practices as material which binds together and stabilizes the thin structure of belief.

I offer a different kind of response to Rorty, one which centers the artifact. What we need is a better method *of shoring up the hull*, of using sheaths of copper and lead and planks of iron and titanium. In this chapter, I want to render accessible these more robust materials by articulating how to incorporate nonlinguistic materials into our epistemic practices. This argument takes a different form from that in the preceding chapter. While the previous chapter articulated the first dimension of artifactual pragmatism by explicating how artifacts compose a *space of practice* between the space of causes and the space of reasons, this chapter takes up how artifacts can be deployed as evidence within communicative and persuasive contexts, that is, as explicit, components of the space of reasons itself.

The chapter will proceed in three parts. In the next section (3.2), I will outline Rorty's ambiguous positions regarding the scope of the conversation and the justificatory role of metaphors. In the following section (3.3), I reframe this ambiguity in terms of the elimination and

rehabilitation of evidence within late 20th century philosophy, focusing extensively on implicit post-Rortyan arguments about photographic knowledge. This takes the form of a broad dialectic—I argue that Donald Davidson's and Robert Brandom's conceptions of *triangulation* on one hand and John McDowell's and Michael Ayers' appeals to *perceptual contents*, on the other, can be combined to allow for assessments of objects as reasons. In the fourth section (3.4), I then argue, using Brandom, that this view opens up the possibility for *artifactual reasons*, where epistemic reliability can be leveraged from within the space of reasons. The resulting position has three major insights—it directly grounds us in the material world of common sense despite our participation in the space of reasons, it allows that very material world to be taken up into our epistemic practices as evidence, and it provides a model for assessing the *reliability* of evidence as a basic component of our commitments and entitlements.

### 3.2 – Slabs, Slaps, and Photographs: The Scope of Conversation

The central issue of this chapter is best described as an attempt to articulate and clarify the scope of conversation. Richard Rorty, in situating himself between analytic epistemologists like Sellars and continental rationalists like Habermas, plays a key role in this clarification. Yet Rorty ultimately rests his case on a loose and shifting definition of conversation. On one hand, Rorty variously appeals to Sellars' *space of reasons* as the model of the conversation, treating the 'conversation' as possessing an explicitly justificatory scope. In the conclusion of *Philosophy and the Mirror of Nature*, he explains that "much of what I have said about epistemology and its possible successors is an attempt to draw some corollaries from Sellars's doctrine" (Rorty 1979, 389). He then goes on to repeat this Sellarsian doctrine as the view that

> in characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says. (Sellars [1956] 1997, 76; quoted in Rorty 1979, 389)

This variation of the conversational ties back to Rorty's earlier defense of "conversational justification" as following from the Quine-Sellars synthesis:

> I shall be arguing that their holism is a product of their commitment to the thesis that justification is not a matter of a special relation between ideas (or words) and objects, but of conversation, of social practice. Conversational justification, so to speak, is naturally holistic, whereas the notion of justification embedded in the epistemological tradition is reductive and atomistic. (Rorty 1979, 170)

If Rorty were to closely follow this Sellars-inspired model of conversational justification, then conversation would be defined in terms of sentences. These sentences would possess terms made meaningful by their inferential role and they would play a justificatory role by serving as reasons for other sentences.

However, Rorty's project is broader than that of Sellars'. At times, Rorty wants to include any kind of performance contributing to the conversation. Even in his more analytic writings, he expounds upon the importance that *abnormal discourse* plays in rupturing and redescribing old vocabularies, suggesting that certain non-justifying linguistic performances help redirect us to new problems beyond what is currently justifiable (Rorty 1979, 320). This leads Rorty to endorse controversial theses on the status of conceptual change and justification. Ultimately, since the units of abnormal discourse merely engender conceptual change, there is no mode of discourse which would count as a reason for or justification of our possession of our current set of concepts, taken as a holistic bundle or vocabulary. Exercising our current concepts to justify the present vocabulary would amount to circularity, and developing new concepts would amount to breaking with current standards of rationality, uprooting them for something else. As a result, Rorty controversially concludes, "All that can be done to explicate 'truth', 'knowledge', 'morality', 'virtue' is to refer us back to the concrete details of the culture in which these terms grew up and developed" (1980, 173). There is no non-circular form of justification which can be offered up as reasons for our current standards and practices.

### *3.2.1 – Three Distinctions: Force-Persuasion, Sentence-Metaphor, Reason-Cause*

I argue that there is an opening here. The 'concrete details' of culture can act as the mortar for stabilizing our practices. But this requires treating these details as components of the conversation in a manner Rorty potentially excludes.

Clarifying Rorty's position on the conversation requires illuminating three interrelated distinctions that he mobilizes throughout his oeuvre: force versus persuasion, sentences versus metaphors, and reasons versus causes. In several respects, the first distinction is one of the central recurring ideas within Rorty's philosophical project, demarcating the difference between abnormal and normal discourse at a theoretical level, and revolutionism versus liberal reformism at a political one. In Rorty's most direct appeal to the force-persuasion distinction, he writes that, "the only

important political distinction in the area is that between the use of force and the use of persuasion" (1989, 84). On the face of it, there is an obviousness to this position—political purges of one's enemies cannot be a substitute for social deliberation, torture cannot replace reasoned argument, and warfare is very rarely preferable to compromise. However, Rorty ends up admitting that this principal distinction is a fuzzy one, that "brainwashing, media hype, and what Marxists call 'false consciousness'" cannot be classified as strictly force or persuasion (1989, 48).

This ambiguity bleeds into Rorty's other distinctions, which are largely holdovers from Donald Davidson's philosophy of language. Davidson contributes to the *sentence-metaphor distinction* with his view that metaphors are non-informational or non-semantic causal impacts from outside of discourse (Davidson 1978). The *reason-cause distinction* then follows from the others.[50] Since sentences are semantic entities, they can meaningfully justify other sentences in a manner that metaphors cannot. And since beliefs are expressed through sentences or are understood as propositional attitudes, Davidson's picture of justification is a coherentist one—only a belief can justify another belief (Davidson 1983).

Rorty follows in Davidson's footsteps by maintaining that the distinction between truth-statements and metaphors is captured by the difference between "familiar and unfamiliar" strings of marks and noises (1989, 17). Just as unfamiliar marks and noises fail to justify, so do non-sentences also fail to be truth-functional since they lack a propositional or sentential structure. Out of convenience, Rorty lumps them in with the metaphors:

> Tossing a metaphor into a conversation is like suddenly breaking off the conversation long enough to make a face, or pulling a photograph out of your pocket and displaying it, or pointing at a feature of the surroundings, or slapping your interlocutor's face, or kissing him. Tossing a metaphor into a text is like using italics, or illustrations, or odd punctuation or formats (1989, 18).

At first glance, this view is perhaps unusual or unexpected. But it follows logically from Quine, Sellars, and Davidson. Even a staunch critic of Rorty such as Jay Rosenberg (1993), while lambasting Rorty's position on metaphor and his loose style of philosophizing, admits that pointings and photographs are not truth-value candidates and never can be (206). Yet when pressed on his treatment of metaphors as incapable of standing as reasons, Rorty slightly wavers:

[50] In my reading, Rorty treats the persuasion-force, sentence-metaphor, and reason-cause distinctions as correlates of each other. In this sense, no formulation of the distinction has explanatory priority over the other. For exegetical purposes, my own argument orders them starting with Rorty's most casual formulation and moving to his most epistemologically complex one.

> Metaphors can justify a belief only in the same metaphorical sense in which one may 'justify' a belief not by citing another belief but by using a non-sentence to stimulate one's interlocutor's sense organs – hoping thereby to cause assent to a sentence. (As when someone holds up a probative photograph and asks 'Now do you believe?') (Rorty 1987b, 169).

On the face of it, Rorty asserts, yet-again, a sharp distinction between sentences and metaphors, reasons and causes. But the distinction is beginning to break down. If sentences are strings of marks and noises, why do they stimulate one's interlocutor's sense organs in a different manner than a photograph would? Or more importantly, if linguistic metaphors may one day be truth-value candidates within normal discourse, what is preventing us from inventing a vocabulary in which the metaphorical sense of 'justify' is made literal? Why can't justification be equally applied to pointings and photographs as with sentences?

If one takes Rorty at his word, sentences appear to slip easily through the skull to directly latch onto beliefs while all non-sentences must first be converted into sentences in order to justify. But why *must* this be the case? Can't I prove *that* I am angry at my interlocutor by slapping them in the face? Can't italics radically change the meaning of a sentence?[51] And don't we consider, according to the typical, non-philosophical, non-metaphorical use of the term 'reason,' photographs—or at the very least, the contents of photographs—to be *reasons* for a belief?

### 3.3 – Rehabilitating Evidence: Triangulation and Perceptual Content

Despite their attempts to naturalize our use of language, the position of Rorty and Davidson looks a lot like the strawman of linguistic idealism. At the very least, there is an uncomfortable divide in their view between the causal world at the outside and the self-contained web of beliefs within ourselves. No matter how much Rorty and Davidson decry Quine's scheme-content distinction, they maintain some form of his linguistic veil. Since sentences can serve as reasons but other types of entities cannot, the sentences begin to take the shape of a scheme which encircles and then shrouds the remainder of the world. At its worst, it looks as if the whole world is a construction of our web of beliefs, that we cannot really get to the world-stuff beyond sentences. Perhaps, we can touch this stuff and see it and hear it. But it can never operate, in thought, as material for justification. It always sits on the sidelines of the game of giving and asking for reasons.

[51] Take for instance, '*She* never said that' versus 'She never said *that*.'

The issue here is one of *evidence* and its elimination. While Rorty and Davidson dismiss Quine's talk of experience, they are comfortable with the remaining epistemological edifice of his project, in particular, his account of the web of belief as an internal system of justification which is only connected to the external world at its edges. Yet even with causal constraint as a substitute for experience, Rorty and Davidson seem to uphold a dualism between the intake and web of belief which re-instantiates the epistemological skepticism they hope to avoid. In other words, calling the intake 'causal' rather than 'experiential' does little to solve the problem. The problem is already present the moment the vocabulary of intentionality—of beliefs, desires, and their justifications—is radically separated from the vocabulary of ordinary physical objects in the world of common sense.

Although Quine was committed, against some of the logical positivists, to reviving talk of ordinary objects over sense data, he self-admittedly ran into problems on how to naturalize sense data. He first considered "observable objects and events" (1992, 2) in line with his physicalism. However, he saw that "a gulf yawns between them and our immediate input from the external world, which is rather the triggering of sensory receptors" (2). To solve this problem, he decided to substitute 'observation sentences' in place of the standard role provided by 'evidence' as observable objects and events (2).[52] His reasoning was that immediate stimulations of our sensory surfaces are different physical events in the world which are not publicly available for assessment by other interlocutors. For instance, A's retinal stimulations *s* by photons bouncing off of object O are one event, B's retinal stimulations *t* by photons bouncing off of object O are another. Quine holds that since the immediate causal influence is the stimulations in A and B and since these are different physical events, the stimulations cannot be the basis for perceptual knowledge. This is because there is no way for A and B to directly perceive O. Instead, they are directly perceiving different sensory stimulations *s* and *t*.

The introduction of language allows A and B to correlate stimulations *s* and *t* such that the basis for conceptual error can occur, and perceptual knowledge can be attained. As a result, observation sentences—such as "There's a rabbit" or "It's raining"— are the basis for the kind of publicity which is necessary for our concepts to be capable of error. Language, quite literally, constructs these objects as posits, or as Quine states, "This is where I see bodies materializing, ontologically speaking: as ideal nodes at the foci of intersecting observation sentences" (1992, 24).

[52] For Quine's earlier account of observation sentences, see Quine ([1960] 2013, §10).

Quine treats the contents of perception in terms of these observation sentences. Since it is actually the case, that 'A perceives *s*' and 'B perceives t' we cannot say that 'A and B perceive O'. We must translate this to 'A and B perceive *that p*' where *p* is some proposition 'x is F' which A asserts when A perceives *s* and B asserts when B perceives *t* (1992, 65). Once this maneuver is set up, Quine's variation of naturalized epistemology proceeds in terms of sentential beliefs or propositional attitudes, like *p.* These sentences can serve in justificatory relations to other sentences. But the world itself, the object O, drops out as a theoretical notion, it does not play any epistemic role in perceptual knowledge, and can only be conceptualized as an object *after* the introduction of language and the coordination of observation sentences.

The position I develop and defend revives the evidential role played by O within Quine's model. I argue that ordinary objects can serve as reasons in the same manner as marks and noises. The idea is that fully rejecting the scheme-content distinction requires rejecting the view that sentences are vehicles of knowledge over-and-above other objects in the world. But, in escaping from Rorty-Davidson coherentism on one side, we find Sellars' Myth of the Given on the other.

Nudging my way to the center, I will now consider various attempts to rehabilitate objectivity from the earlier Rorty-Davidson coherentism. Following Steven Levine's (2019) example, I separate these proposals along *communicative-theoretic* and *experiential-theoretic* or *perceptual-theoretic* lines. On the communicative-theoretic side, I look to Donald Davidson's and Robert Brandom's respective semantic externalist revisions to coherentism which challenge Quine's elimination of O causally but not epistemically. On the experiential-theoretic side, I draw from proposals to extend reasons beyond sentences—John McDowell's insistence that perception already includes conceptual capacities and Michael Ayers' position that nonconceptual contents of perception can justify without falling into the Myth of the Given. I will ultimately reject these theories as suitable responses to the problematic at hand, but I take up an element of each in order to construct my version of an artifactually-extended communicative-theoretic position.

### *3.3.1 – Triangulation, Externalism, and Evidence*

In his "A Coherence Theory of Truth and Knowledge," Davidson follows Quine's rejection of evidence as a theoretical notion, writing that the philosopher should "give up the idea that meaning or knowledge is grounded on something that counts as an ultimate source of evidence" (Davidson 1983, 145-146). But by his later work, Davidson is increasingly troubled by Quine's elimination

of evidence and its bearing upon his own epistemological theory. He stresses a tension between Quine's "essentially first person and Cartesian" theory of knowledge and his quasi-behavioral naturalized epistemology (Davidson 1990a, 194). Davidson thinks Quine's best bet of resolving this tension is to reject the former for the latter. Naturalized epistemology would then be free from the last vestiges of empiricism and take its starting point from an essentially third-person standpoint of human beings interacting with public objects (Davidson 1990b, 47-48).

Davidson's defense of this position depends upon shifting the objects of perception from the immediate sensory stimulations (what I have termed *s* and *t*) to the public objects that cause these stimulations (what I have called O), or from the *proximal stimulus* of perceptual beliefs to the *distal stimulus* (1990b, 54). Against Quine's skeptical-sounding conclusions, this model follows the standard set by Wittgenstein and Dewey by describing one's learning a language in terms of ostensive learning, training, or habit modification rather than from a first-personal standpoint of shifting perceptual states or immediate sensory stimulations.[53]

Davidson motivates the shift from the proximal to distal stimulus through the concept of *triangulation*. What Davidson points out is that an externalist account of language acquisition would document instances of a "triangular arrangement of the two agents and a commonly observed object" (1990b, 50). He later elaborates, "Person A responds to person B's responses to situations both A and B find similar. A triangle is thus set up, the three corners being A, B, and the objects, events, or situations to which they mutually respond" (1997a, 140). Bringing back Quine's model, Davidson thinks the social dimension imparted by triangulation picks out the shared object of perception, essentially bypassing sensory stimulations *s* and *t*. The basic principle at play is publicity—A cannot respond to B's simulations *t,* nor can B respond to A's simulations *s*. But if A and B are to treat each other as responsible agents, as following a rule, they must understand themselves as situated within the same world. The way to get at this shared world is to trace the causal history of A and B's perceptual beliefs beyond the immediate proximal stimulus. When we do this, we find that the only available *shared cause* of their beliefs is O, the distal stimulus, the triangulated object. Essentially, sociality is the dimension which allows complex organisms to pick out shared objects in the world.

[53] Davidson points out that Quine sometimes refers to a distal stimulus rather than a proximal one. For the sake of argument, I have picked out Quine's references to the latter. I have also included as a reference above Quine's (1992) later endorsement of the proximal stimulus in response to Davidson in *The Pursuit of Truth*.

It is perhaps easiest to see this with the phenomenon of hallucination. When A and B both correctly perceive O, A's perception is caused by proximal stimulus *s* which is caused by distal stimulus O and B's perception is caused by proximal stimulus *t* which is caused by distal stimulus O. However, if A hallucinates, they will experience a stimulation *s* which lacks a corresponding object O. The rest of the social group will fail to have a similar experience, and will assess A's perceptions to be unreliable. This check on perceptual reliability requires organisms responding to each other and responding to the shared distal stimulus. When the distal stimulus is removed as an explanatory variable, all organisms are responding to their own private stimulations, and the difference between hallucination and veridical perception cannot be made explicit until the introduction of observation sentences to correlate these private stimulations.

But Davidson points out that non-linguistic organisms react as if they are in a shared public world despite not speaking in terms of observation sentences. A dog, in a game of catch, responds *to the thrown ball*, not to its private stimulations. As Davidson explains,

> One sees this in its simplest form in a school of fish, where each fish reacts almost instantaneously to the motions of the others. This is apparently a reaction that is wired in. A learned reaction can be observed in certain monkeys which make three distinguishable sounds depending on whether they see a snake, an eagle, or a lion approaching; the other monkeys, perhaps without seeing the threat themselves, react to the warning sounds in ways appropriate to the different dangers, by climbing trees, running, or hiding. (1997b, 128)

It would be highly unusual and methodologically counterproductive to explain these phenomena according to each individual organism's perceptual and cognitive components, to say that they react to their proximal stimulations. Rather we say that the fish *react to each other* and that the monkeys *react to the lion*. We grant that non-linguistic animals can triangulate the distal stimulus as well.

Despite granting triangulation to animals, Davidson wants to separate triangulation from the possession of beliefs, desires, or judgments. He says that triangulation is "a prelinguistic, precognitive situation which seems to me to constitute a necessary condition for thought and language, a condition that can exist independent of thought, and can therefore precede it" (1997b, 128). In the tradition of Sellars, later furthered by Brandom, Davidson separates "the power to discriminate" from the "possession of the corresponding concept" (1992, 160). Concepts are the components of judgments and judgments are truth-candidates. Because of this Davidson holds that "it is only with concepts and judgments that we can be said to have the idea of an objective world,

a world that is independent of our sensations or experience" (1992, 160). But there is a tension here between the *objects of triangulation* and *concepts*. Davidson writes that a child "thinks something is red or a ball only if it is in some sense aware that a mistake is possible. The child is classifying things, and it knows it may have put something in the wrong slot" (2000, 14). Davidson calls this "the step from reacting to proximal stimuli to the thought of distal objects" (2000, 14). And he adds that this is "where the concept of truth enters," where the possibility of error emerges (2000, 14). But if is the case, then the child in this example would possess concepts prior to knowing a language. Playing a game where a red ball is placed inside the circular hole of a toybox would be sufficient for conceptual understanding and the possibility of error.

Robert Brandom (1994) identifies this tension and resolves it by challenging the idea that triangulated objects can exist without the corresponding concepts. His basic assertion is that triangulation cannot specify the correct object O without supplementation by concepts—for instance, two correlated thermometers do not respond to 'temperature' but to the proximal air molecules surrounding them. For Brandom, an inferentialist semantics is required for adequate corroboration of public objects; *causal triangulation* must give way to *inferential triangulation,* the distal stimulus becoming a product of learning a language (430).

Brandom's proposal satisfies the tension, but I suggest taking the alternative route which treats rule-governed or normative practices as category which extends beyond linguistic norms. Following Allen's (2004) suggestion that norms are present within the use of artifacts, I grant a form of non-linguistic conceptuality to triangulated objects within norm-governed practices. This would allow us to say that, when a dog is adequately operating within social practices, she does react *to the ball* without necessarily saying that she is a full-blooded concept-mongerer who lobs truth-evaluable claims back and forth to others. Language would then operate as Davidson wants it to, as something which expands the basic triangular arrangement; it "fills in and enriches the base of the triangle" (1992, 161) but it does not construct it in the first place. Instead, pre-linguistic triangulation explains how social animals interact with external parts of the environment, parts which become common objects and their associated norms through social interaction and sanctioning. Triangulation, then, is the theoretical mechanism which explains how artifactual normativity (Chapter 2) operates prior to and outside of discursive norms.

*3.3.2 – Conceptual Contents of Perception*

In several respects, the position I am developing comes close to various Deweyan suggestions that a Rorty-Davidson account of meaning fails to capture the norms and standards of common sense. Among contemporary adherents of classical pragmatism, Mark Johnson (2008) stresses Rorty's undeveloped theory of conceptual change, pointing out that "marks and noises *somehow* get us searching for new vocabularies in which they are no longer unfamiliar, but he [Rorty] has no account whatsoever of how this process is supposed to work" (198). Thomas Alexander (2013) has similarly taken aim at the thesis that all our nonlinguistic practices would be regarded by Rorty and Davidson as meaningless. Responding to Rorty's insistence that slaps, kisses, and pointings are "metaphors," Alexander (2013) holds that, "the kinds of actions mentioned—kissing, pointing, slapping—are meaningful parts of any communicative activity, unless, of course, they are performed by a madman" (197).

Analytical pragmatists equally criticize Rorty's narrow linguisticism. Colin Koopman (2011) challenges the linguistic turn as an effective social and political device for Rorty's anti-authoritarianism. For him, the appeal to linguistic practice fails to capture the normativity of our "somatic practices" which do not involve the trading of marks and noises (75). Steven Levine (2019) equally emphasizes somatic practices which are not reducible to linguistic performances. Leveraging the space of reasons in relation to Dewey's principle of continuity, Levine holds that the material of the space of reasons is continuous with the material of our experience (223). In this manner, the space of reasons extends outwards and into our embodied practices.

These criticisms match my own from the previous chapter that Rorty excludes from his position a *space of practice* which is neither causal nor explicitly discursive. Yet as an *epistemological* criticism, these positions are missing an important step. What is needed, which the second dimension of artifactual pragmatism provides, is an account of how normative-yet-nonlinguistic practices figure *epistemically* within the critique of the Davidson-Rorty position. This is to say that there is a difference between (i) asserting the existence of a space of practice or a realm of meaning beyond the verbal and (ii) asserting that the space of reasons can *expand into* and *draw from* this newly posited space to undercut the narrowly linguistic variety of conversationalism.

The above criticisms emphasize that Rorty obscures aesthetic, embodied, or technical modes of understanding, that is, they assert (i), but at the same time, this assertion often acts like

an epistemological criticism or like an instance of (ii). However, insofar as the expanded conception of meaning provided by (i) is supposed to circumvent the conflation of meaning with rationality or epistemology for the classical Deweyans, I am not sure how (ii) is to be argued from (i). This is because the contents of (i) are emphasized as those which do *not* play an epistemic role in our reasoning. In response, the classical Deweyan must argue that the meaningful units posited by (i) can sometimes but not always satisfy the conditions needed for (ii), but here, they run into a second problem. This is one they share with the analytic pragmatists. Even if the contents posited within the expanded space of practice can play an epistemic role in the space of reasons, there is still the question of operation. How do meaningful, nonverbal elements directly justify beliefs in a manner that constrains the conversation from outside?

Levine (2019) points to McDowell as a potential resource here. In his book *Mind and World*, McDowell (1994) provided a famous rebuttal to the Rorty-Davidson account. His central concern was to rescue *experience* from Davidson's insistence that our beliefs are directly caused by the world. McDowell's worry is that Davidson and Rorty present a picture of rationality in which the world does not adequately constrain our beliefs from the outside, since they see experience as beyond the scope of thought, as something outside the system of beliefs. At the same time, McDowell is uncomfortable with appeals to the Myth of the Given, since these assertions contradictorily assume that an empirical intake can provide outside justification for our actions without having the form of a belief. The Rorty-Davidson position rightfully jettisons the idea of the Given, but without rehabilitating experience, McDowell thinks they do not sufficiently explain how we could ever really get outside our system of beliefs to see the world rightly.

McDowell's solution to this dilemma is to assert that experience already has a conceptual content, *that things are thus and so,* enabling it to stand in the space of reasons (26). By possessing this conceptual structure, a kind of propositional content, presentations of the world in experience contain elements of the spontaneity of our thinking in addition to our receptivity to the world. These presentations are structured in a manner available to thought but are also imposed upon us in a manner which is beyond our rational control. Since experience possesses both this spontaneity and receptivity, McDowell asserts that experience "enables the layout of reality itself to exert a rational influence on what a subject thinks" (26). Experience mediates the structure of the world with the structure of thought.

Essential to this view is the controversial thesis that the layout of reality matches the structure of perception and thought. This is the claim that reality is inherently conceptual. As McDowell explains, "*That things are thus and so* is the conceptual content of an experience, but if the subject of the experience is not misled, that very same thing, *that things are thus and so*, is also a perceptible fact, an aspect of the perceptible world" (26). This leads McDowell to assert, contrary to Rorty and Davidson, that the space of conceptuality has no boundary, that the way judgments are configured matches the way experience presents facts and that veridical experiences give the facts of the world.

Despite the unusual conclusion, there is much to like in McDowell's picture. The model of the "unboundedness of the conceptual" suggests that we should reject a certain skeptical notion of the sideways-on picture of interpretation, one which sees an interpreter's inner conceptual system as a resource for interpretation of some other interpretants outside of this conceptual system (as with Quine's radical translation and Davidson's radical interpretation) (1994, 35). McDowell helpfully modifies this account by seeing two persons, an interpreter A and interpretant B as already *inhabiting* the space of reasons together:

> We are not filling in blanks in a pre-existing sideways-on picture of how her [B's] thought bears on the world, but coming to share with her a standpoint *within* a system of concepts, a standpoint from which we can join her in directing a shared attention at the world, without needing to break out through a boundary that encloses the system of concepts. (1994, 35–36).

McDowell further elaborates that an upshot of his view is that, in contrast to accounts of non-conceptual experience as a kind of embodied coping, he rejects the notion of a disembodied intellect, a mind of beliefs and desires which stands above a body engaged in practice (2007, 349).

In rejecting non-conceptual experience, McDowell's target here is Hubert Dreyfus, who accuses McDowell of mentalism on the grounds that we often interact with affordances from the world without possessing beliefs or justifications for these actions (Dreyfus 2005, 59).[54] Dreyfus wants to say that McDowell's project is too Kantian, too intellectual. To this, McDowell gives a sobering response, one essential for my own position. He explains that his model of rationality *transforms* prior embodied habits and affordances into potential reasons. He writes, "Affordances

[54] In the prior chapter, I have agreed with Dreyfus that there is a difference between the space of practice and the space of reasons. However, I took the normativity of practice to imply that skillful coping already possesses concepts (or ways of ascertaining correctness and incorrectness) even when the norms cannot be made explicit.

are no longer merely input to a human animal's natural motivational tendencies: now they are data for her rationality" (2007, 344). McDowell's point is not to deny the existence of embodied practices but to claim that there is no need to exclude the skills and materials of embodied practice from the broader practice of giving and asking for reasons—for McDowell, non-propositional or non-linguistic components of experience can be reasons.

At the same time, Dreyfus's challenge is correct insofar as McDowell's position contains a Kantian and Hegelian baggage that often inflates, rather than deflates, the philosophical problems at hand. The Kantian language of spontaneity and receptivity has the potential to reinstate an anti-naturalist, a priori model of philosophizing and the Hegelian language of unbounded conceptuality risks treating the world as possessing an innate logical structure, as being ontologically constituted by the thinkable.

This is also Michael Ayers' (2004) challenge to McDowell. Taking up the example of a photograph, Ayers worries that neither Davidson nor McDowell can endorse the common appeal to photographs as reasons or as evidence. He gives the example of two interlocutors discussing the assassination of JFK:

> Suppose it is asked, 'Does this photograph justify the belief that someone was shooting at President Kennedy from the ground?' Obviously we need all kinds of beliefs about the photograph before we get to ask this question, beliefs about who took it and where, where it has been since the event, whether it has been tampered, and so forth. Suppose we rightly have the appropriate beliefs about its credentials, and are now focusing on what the photograph presents (or re-presents), the events recorded, the photograph's visual content. (Ayers 2004, 243–244).

Ayers first challenges the Rorty-Davidson position that the photograph "cannot justify *any* conclusion about the events photographed, since only a belief can justify another belief, and a photograph is not a belief" (244). Ayers alternatively wants to say that the "the photo-content justifies or grounds both the judgement about the photographed state of affairs and the judgment about the photograph" or in other terms, that "what is asserted about the world may be justified by what a photograph *shows* about the world (244, my emphasis).

Perhaps McDowell's position hints at this latter view, but Ayers (2004) does not think so. He thinks that McDowell has fallen into a similar trap, that his "talk of facts is just too obviously a projection of propositional form onto reality" (248–249). Defending non-conceptual contents of experience, Ayers points out that descriptions of some material object will be propositional, but

the object itself or any pictorial representation will not be propositional and thus will not be conceptual (249). He therefore accuses McDowell of the view that perception is quasi-linguistic.

The straightforwardness of Ayers' approach is useful, as Ayers' wants to say that photographs can be reasons for a belief, full stop. Yet Ayers' conclusion, as articulated, seems to fall prey to the Myth of the Given. To take up a Sellarsian view of the matter, Ayers seems committed to saying that *seeing* these objects amounts to *knowing* them. But this is surely not the case—young children will not see the contents of the photograph of JFK as evidence. Similarly, if one were to send a photograph back in time to medieval Europe, neither the lords nor the landed peasantry would see the photograph *as* a photograph, as something bearing an evidential function which is different from paintings or drawings. The ability to see the contents of the photograph as representing the objects depicted and to see photographs *qua* photograph (bearing in mind its evidential function) are learned abilities which require some conceptual background. The Sellarsian point is that photographs must first be integrated into a broader rational system in order to serve as evidence.

But Ayers is on the right track, that it is possible to appeal to the contents of a photograph *qua* photograph as reasons for a belief. However, I think this requires extending the component parts of the space of reasons outward to involve nonlinguistic units rather than relying on the non-conceptual contents of perception.

This is precisely what McDowell does in his response to Ayers. He explains that his notion of 'conceptual' should not be confused with propositional expression.[55] In a retort, he states that "Ayers is disallowing the very possibility of supposing, as I do, that a conceptual mode of presentation might itself be a sensory mode of presentation" (McDowell 2006, 135). McDowell accordingly revises his position to clarify that perceptions are not linguistically structured. Rather, the claim that perceptions are conceptual is minimally the claim that they are so structured so that one *could* refer back to them as reasons within instances of reflection (128–129). He gives the example of a person following the directions of a signpost:

> Consider someone following a marked trail, who at a crossing of paths goes to the right in response to a signpost pointing that way. It would be absurd to say that for going to the right to be a rational response to the signpost, it must issue from the subject's making an explicit determination that the way the signpost points gives her a reason for going to the

[55] McDowell perfectly expresses the confusion in the following passage: "I proposed to modify Davidson's slogan by saying that not only beliefs but also experiences can be reasons for belief. And according to my old assumption experiences have the same kind of content as beliefs." (McDowell 2008)

> right. What matters is just that she acts as she does because (this is a reason-introducing "because") the signpost points to the right. (This explanation competes with, for instance, supposing she goes to the right at random, without noticing the signpost, or noticing it but not understanding it.) What shows that she goes to the right in rational response to the way the signpost points might be just that she can afterwards answer the question why she went to the right—a request for her reason for doing that—by saying "There was a signpost pointing to the right". She need not have adverted to that reason and decided on that basis to go to the right. (McDowell 2006, 129)

With this example, what McDowell shows is that perception affords us the material necessary to be able to retroactively make such claims, and therefore, that conceptual and rational capacities are "operative also in the unreflective response" because that response is produced from and available for use by rational claims (130).

With this new model of how reasoning relates to experience, McDowell now provides his own analysis of photographic evidence in defense of his and Davidson's shared Sellarsian heritage:

> Certainly one can adduce a photograph as evidence for a belief about the assassination, not just out loud, trying to persuade someone else, but also in deciding what to think. The point of Davidson's slogan is not to deny that, but to claim that the photograph can play that role, a role in an operation of rationality, only by figuring in the content of an actualization of capacities that belong to one's rationality: that is, capacities that are conceptual in the sense of my stipulation." (McDowell 2006, 138)

In other words, McDowell thinks that his model correctly allows us to figure the contents of photographs as reasons for a belief by appealing to the manner in which they are conceptual under his new definition—the contents of photographs are those which can be appealed to by claims within the space of reasons.

This comes close to rehabilitating evidence, but I am not so sure McDowell adequately solves the problem, since he still relies upon Kantian first-personal and subject-centered language, where the *perceptions* and not the photographs are the things which have content and which perform evidential work. Even when trading out his initial references to propositional content in experience (McDowell 2008, 258), McDowell's later work still mobilizes Kant's appeal to *intuition* (260), where the separation of intuition from judgment allows him to disentangle perceptual contents from propositional claims (264). McDowell's later point is that intuitions are conceptual but non-propositional. They are of the same conceptual kind as judgments but cannot be expressed propositionally in the same manner as judgments (262). So, McDowell insists that we should not think of experience as propositional or discursive in scope, since it does not consist

of propositions as judgments do, but experience can still be structured by conceptuality in terms of its being an intuition.

At the risk of sounding dismissive, I do not see why this Kantian vocabulary needs to be revitalized. McDowell's Kantianism resists a naturalistic account of how we experience the world, trading contemporary biology and cognitive science for a pre-personal philosophical psychology of sorts. The issue is that McDowell resists an externalism whereby the contents of perception are traced back to triangulated public objects. As Davidson explains, in criticizing the need to revitalize experience, "perceptual sentences have an empirical content given by the situations which stir us to accept or reject them, and the same goes for the beliefs expressed by those sentences" (Davidson 1997a, 137). In this manner, a Davidsonian picture questions whether there is a need to invoke experience and its contents at all. Or at the very least, it questions the notion of a vocabulary which describes various kinds of anti-naturalistic, non-public objects such as the "contents of intuition." What is needed instead is a reconstructive account of how the normative realm of *practice* directly participates in the game of giving and asking for reasons, or an account of how the *objects* of practice can symbolize, represent, critique, negotiate, and persuade.

### *3.3.3 – Reasons and Objects*

My own position takes up elements of both the communicative and experiential models. I adopt Davidson's triangulation and his thesis that the distal stimulus is the causally relevant object of our thoughts. I likewise adopt McDowell's thesis that extends reasons beyond belief into other domains where conceptual capacities are operative. But endorsing these theses requires rejecting others. In holding McDowell's thesis on reasoning, I ultimately reject the Davidson-Rorty position that only a belief can serve as a reason for another belief. Yet in my endorsement of Davidson's externalist picture of semantic content, I reject McDowell's empiricism and his Kantian appeals to intuition. For Davidson, two individuals can *act with* and *talk about* the same objects in the world. For McDowell, even later on, the notion of "perceptual content" or "intuition" gets in the way of this acting with and talking about. It reenforces a return to the individual subject as a matter of concern, which then requires a "picturing" relation to be asserted after the fact in order to ground us in the rational structure of the world.

Even if McDowell is able to recast his idealism into a form of "common sense," his position of the "unboundedness of the conceptual" falls into the same pitfalls as Brandom's position

outlined in the previous chapter, namely that conceptuality cannot be substantively asserted as a special structure to reality prior to human thought and action. Avoiding this move requires the Rorty-Davidson separation of the space of causes from the space of reasons or the view that "rational relations" only exist as emergent features of social practice rather than as facts or states of affairs already in the world prior to thought.

The picture I endorse involves an account of how triangulated public objects can stand-in as reasons. To differentiate this position, we can return to the example of an individual turning right at a signpost on a marked trail. Here, the Rorty-Davidson position looks *sideways-on* at the situation. They see the individual behaviorally reacting to a signpost. This reaction is a *causal relation* between the object and the individual. The signpost causes a perceptual belief, and this perceptual belief then justifies other relevant beliefs leading to the individual's going right. We should note that the rational apparatus of the total web of beliefs is internal to the individual, and this apparatus is merely causally connected to the outside distal stimulus, the signpost, which itself does not justify.

McDowell takes up a different approach. From a *first-personal* perspective, he describes how a rational individual possesses the capacity to refer back to their perceptions as reasons. So, an individual has an experience of the signpost which is structured in such a way that they can later claim "I went to the right *because* the signpost pointed right." The perceptual experience of the signpost serves in a justificatory, rational relation to the belief expressed by this claim. According to McDowell, contra Rorty and Davidson, experience can ground beliefs after all, since it possesses its own conceptual structure which, though not propositional, can be expressed through propositional contents.

Taking hold of these two positions, I argue for a third possibility. Imagine a scenario in which Person A watches as Person B walks along the marked trail. A sees B pass upon a signpost which points right. B pauses in front of the signpost, glances left and right and then continues walking to the right. Later, A watches a squirrel walk along the same path. The squirrel also stops before the signpost, glances left and right, and then continues rightward. My position is as follows. So long as A can infer that B understands the *meaning* of the signpost, A should attribute the signpost as the *reason* for B's going right. However, A should not attribute anything other than a causal relation between the squirrel's behavior and the signpost.

Since the squirrel (or we could also consider a very young child) is not trained to understand the signpost as anything other than an object presented against its background, it cannot take up the signpost as a reason for going right. The signpost can only be used as a reason by those possessing the relevant training, those who see it, for instance, as an artifact bearing some official status that demarcates the boundaries of a trail. Merely seeing the signpost as some object against its background would not count as utilizing it as a *reason* because the object would not be understood to have a socially and epistemically relevant role in a practice. As such, the signpost does not act as an epistemic given nor does A's attribution of the signpost as a reason for B's actions require conflating reasons with causes.

At the same time, A and B interact with the *signpost* rather than proximal sensory stimulations of the signpost. Following Davidson, we should say that B's beliefs are related to the triangulated distal stimulus itself. It is not B's perceptual capacities which render the signpost a reason, but the conceptual structure of the social practice in which both A and B are situated. Here, A attributes proper training to B, while B possesses the same capacity to make such attributions upon A. This is to say that both A and B understand themselves as governed by the implicit norms of their socially constituted environment. This mutual recognition of each other and of the signpost requires that the signpost be a public object whose status can be tracked by both A and B. Neither A nor B's perceptual states are alone sufficient to derive this social-normative structure. Nor can a Quinean model of stimulus synonymy track the rational and epistemic role of the signpost as an object in the world. Rather, the contents of A and B's perceptions must be understood as the triangulated contents of the public world itself. It is the world which is normatively transformed through practice and persuasively operative through discourse.

However, artifactual pragmatism departs from the prior post-Sellarsian positions outlined in this section (3.3) because it defends the position that non-linguistic practices can be normative. The explication of the space of practice in the preceding chapter emphasized how certain norms can be established through social sanctioning without these norms being made explicit in language. There I defended an account of *practice* as an intelligent, normative, and conceptual activity, but an activity which nonetheless remains distinct from *reason* in that it is not persuasive or

justificatory. The important difference is that practice involves agents acting according to norms of correctness and incorrectness but not according to norms of truth and falsity.[56]

In making this distinction between reason and practice, my inclusion of the signpost into the space of reasons does not undermine the centrality of the discursive and its connection to rationality. In many respects, my position closely follows McDowell in that I merely claim that discursive capacities enable non-discursive contents to be utilized in the space of reasons. This is in contrast to a model of discourse which excludes non-linguistic entities from justificatory practices (Rorty and Davidson) and from an idealism which "propositionalizes" all non-discursive entities into states of affairs (Brandom). My point is that the space of reasons can *require* language as a necessary component for ascribing rational capacities to persons without the space of reasons being entirely *constituted by* language. The main difference between artifactual pragmatism and McDowell's view, then, is that I defend a social and externalist mechanism for conceptualizing non-discursive content—the mechanism of social practices given in the preceding chapter—while McDowell defends a Kantian mechanism in which the non-discursive content of our perceptual intuitions is already conceptual.

Since I do not intend to conflate all practices with the space of reasons, Rorty's force-persuasion distinction still holds. It distinguishes between moves internal to the space of reasons and all outside performances. However, now the same type of performance may count as either force or persuasion depending upon the context. Take the instance of pointing at an object O. Pointings justify as an effort to coordinate social triangulation with background concepts. In most instances, when A points out an object to B, A is not merely causally cutting through the air with their outstretched hand or bouncing new streams of photons off of their skin and onto B's retinas. A is *intentionally* directing B's attention to some object in an attempt to recenter the bounds of discourse through triangulation. A is using a triangulable object O to generate a belief in B.[57] This is an intentional (practical) and persuasive (rational) performance, since we are to assume B

[56] In my configuration of the three "spaces," each space is nested within the other. All instances of reasoning are instances of practice, and all instances of practice take place in the causal world. One way to parse it out is that non-intentional actions are causal, causal actions which are also intentional are practical, and practical actions which are also truth-involving are rational.

[57] And from the above elaboration of triangulation, we should question how the use of strings of marks and noises is any different than this instance of using an object O to generate a belief *p* in *B.* Sentences, according to artifactual pragmatism, are just new, compound objects formed from past materials. To persuade with a sentence is just another instance of using a triangulable object to generate a belief in one's interlocutor.

already has the requisite conceptual background to attain the belief *p* from O which A intends them to have. When B does not have this full discursive background, A's pointing is a practical association between objects, gestures, and sounds. It is an instance of what Wittgenstein calls *ostensive teaching* ([1953] 2009, §6), but it is not an instance of persuasion or reason-giving.

Where Dewey and Wittgenstein depart from the later neo-pragmatists is that they hold that, once a discursive practice is set-up, all its constituent parts—including such things as slabs, samples, and pointings—are available as *reasons*. In *The Blue Book,* Wittgenstein gives a clear example of how this can be argued:

> Suppose I pointed to a piece of paper and said, to someone: "this colour I call 'red'". Afterwards I give him the order: "now paint me a red patch". I then ask him: "why, in carrying out my order, did you paint just this colour?" His answer could then be: "This colour (pointing to the sample which I have given him) was called red; and the patch I have painted has, as you see, the colour of the sample". He has now given me a reason for carrying out the order in the way he did. Giving a reason for something one did or said means showing a *way* which leads to this action. In some cases it means telling the way which one has gone oneself; in others it means describing a way which leads there and is in accordance with certain accepted rules (Wittgenstein 1958, 14)

The important take-away of the passage is that Wittgenstein clearly delineates the act of ostensive teaching from the painter's later justification of his actions. It is also essential that this justification involves *showing a way*—it cannot be fully performed without the color sample and the wall standing in as triangulated objects which have been taken up into the act of justifying. At the same time, Wittgenstein writes that if the painter had merely provided a description of the process of training and of his habitual responsiveness to it, he would only have given a *causal* account and not a rational one (15). This leads Wittgenstein to deduce that the chain of causes leads to an infinite regress but that "the chain of *actual* reasons has a beginning" (15). Where Wittgenstein departs from the later neo-pragmatists is that he posits a beginning to reasoning by leveraging the triangulated objects as components that are internal to the game of giving and asking for reasons. The difference is that, though Davidson provides a roughly-Wittgensteinian account of how our triangulation with public objects causally leads to the instantiation of language, the triangulated objects drop out of the resulting rational edifice. For Davidson, we *talk about* objects. For Wittgenstein, the possibility of *talking about* objects also enables the possibility of *talking with* them—words are just one of the many tools of the language through which we can *show a way*.

### 3.4 –Artifactual Reasons

In understanding justification as *showing a way*, the space of reasons becomes populated by all of the ordinary objects which make up the unphilosophical world of common sense. However, though pointing might leverage any object as a reason, it is essential to see that Wittgenstein's case of the color sample departs from mere instances of pointing. For the color sample to serve as a reason, the two persons must not only understand the meaning of the words but also the corresponding meaning of the *vehicles of expression*, the very material in which abstract concepts are being deployed. Otherwise, the red color sample would perform the same function in the space of reasons as the act of pointing to the color of the wall and saying, "But this is red!" or pulling out a red colored marker from one's pocket. This is to say that the sample is more than a reminder of the color.[58] Wittgenstein's position requires understanding the conceptual role played by the artifact as it grounds the use of words within our language or as it is embedded within a prior system of commitments and entitlements. Our triangulations include *artifacts* in addition to their contents. They are the objects through which background practical know-how becomes entangled with and mobilized within reasoning.

This is just the insight needed to properly understand the status of photographs in the space of reasons. To recap, Rorty takes up the photograph as a metaphor, a slap in the face, but one which nonetheless might contribute to "the conversation" in its broadest sense. Striving for some kind of objective constraint, McDowell treats the photograph as conceptual and rational, but only insofar as it can be mobilized through our conceptualized intuitions. I have suggested (in 3.3) that neither gives us the necessary explanatory framework to seriously consider the epistemic role of evidence.

Instead, I want to claim that triangulation provides a model for how different interlocutors conceptualize the contents of visual evidence and that this model is more complex and socially-embedded than prior analyses have suggested. Davidson perhaps paves the way to this approach, when he, in describing the appearance of his own writings in artworks by Robert Morris, takes up triangulation as a model for explaining how the contents of the visual arts are embedded in a system of meaning. Davidson explains, "[Morris] has put his viewers in a position to triangulate with him the location of his creative acts" (1992, 165). While Davidson points out that, the paintings'

[58] I see something contractual in the handling of the color sample. It seems like what separates a rational from a causal explanation in Wittgenstein's view is the way that the sample is treated as itself expressing a binding obligation among the participants.

"coordinates" are especially particular in their scope and context, he is also moved by the suggestion that his own work may be used to clarify this "larger canvas" (1992, 165). Triangulation here is mobilized as a means to understand the way that discriminative perceptual responses and shared conceptual backgrounds interact with non-linguistic artworks. But despite this radical shift in perspective, Davidson never gives a deeper notion of the intricacies of how art should function in reasoning. For all his rehabilitation of public objects, these objects still fail to operate as evidence.

### *3.4.1 – Reliability and Pictorial Concepts*

The broader effects of the elimination of evidence have led to the dismissal of instruments and non-linguistic vehicles of expression, such as photographs or images, as unique modes of justification deserving of serious attention. Most epistemological views on the workings of artifacts take up a subject-centered analysis where artifacts are of secondary importance, a mere corollary of perceptual or testimonial knowledge.[59] Instead, I consider signposts, thermometers, photographs, and other kinds of artifacts to operate as *artifactual reasons*. This requires shifting our focus from the individual subject as the central loci of knowledge to the artifact as a social object. Photographic knowledge depends less on cognitive, perceptual, or linguistic explanations, and more on the epistemically relevant properties of photographs as they are constituted within the social practices involved in their production, regulation, and distribution. This is true for all types of artifacts—a mercury thermometer, for instance, is epistemically reliable due to its physical and chemical makeup in addition to standards in thermometer production and in standardized numerical representations of temperature.[60] To *read* thermometers requires understanding the practical know-how of associating the mercury line with the numerical representations and, in many cases, understanding the relation of these representations with various corresponding temperature samples.

[59] To briefly outline a range of views, there is the position that photographs are *transparent* or directly reveal their contents (Walton 1984), that they causally link back to the objects depicted (Scruton 1981), and that they are a form of non-inferential perceptual knowledge (Cavedon-Taylor 2013).

[60] I am inspired here by Habgood-Coote's (2023) comparison of photographic depictions to thermometers, using a theory of instrument-based beliefs (Goldberg 2020). I will lay out Habgood-Coote's analysis in Chapter 5. Hasok Chang's (2007) analysis of the thermometer and the practice of "thermometry" also provides a historically grounded pragmatist account of thermometers as epistemic instruments.

Robert Brandom (1994) is correct to think of these associations within our thinking as inferential links between reliable responses to objects in the world, but he is hesitant to take up a position which directly tracks the epistemic role played by individual artifacts. In his comparison of instruments to human perceptual states, Brandom locates instruments between thought—as a complex inferential network—and the blind dispositions of physical systems.[61] While a thermometer is certainly unable to perform the inferential moves necessary for us to treat its responses to temperature as 'conceptual' in Brandom's terminology, it is nonetheless an artifact for which we have taken great care to locate epistemically relevant dispositions and develop social practices for recording and standardizing these dispositions. What separates instruments from other phenomena in the world is how our social practices index reliable dispositions in order for their outputs to count as meaningful expressions within reasoning. Brandom states:

> Artificial instruments differ from other physical systems, such as chunks of iron, only in having been constructed so that some subset of the partition of possible stimuli into equivalence classes according to the distinguishable responses the instruments are disposed to produce corresponds to some distinction of practical or theoretical significance to the user, who thereby attaches some significance to them. (1994, 88)

Brandom's point hints at a broader project to adapt *reliabilist* theories of justification as a resource for his social pragmatist position.

Reliabilism, a leading theory in analytic epistemology, takes some belief to be justified if it was produced by a reliable process (Goldman 1976, 1979). Reliabilism is considered an *externalist* theory of justification in that it allows an individual to possess knowledge without being able to state reasons for that knowledge. Brandom agrees with the externalist insight that someone could possess knowledge due to a reliable process without themselves being able to provide reasons, but he criticizes reliabilism's theory-neutral standpoint (Brandom 1998). For Brandom, physical systems reliably respond to innumerable other events and objects in the world—a weather thermometer reliably responds to temperature but also can respond with some reliability to sunlight and seasons. Merely asserting that there are systems with reliable dispositions is not enough to leverage those systems as epistemic resources. For that, the relevant dispositions must be indexed by intentional actors in a social practice. As such, we can only appeal to reliability after we have

[61] See Armstrong (1973) for an earlier account of the distinction of dispositions from beliefs.

set up practices of assessment and an edifice of reason-giving which allows for the reliable responses of to qualify as reasons.[62]

Brandom's analysis demonstrates the difference between what I have called *pointings* and *artifactual reasons*. Pointings are attempts to reorient discourse through triangulation and can be enacted with pretty much any ordinary object O so long as practitioners have the relevant concepts for that object and its properties. On the other hand, artifactual reasons involve social practices where some reliably responsive disposition of a physical system has been picked out and appropriately indexed by representational resources. There is a difference then, in pointing to a tree as evidence for the claim 'The leaves are changing' and pointing to a thermometer as evidence for 'It is 30F outside.' The latter class of performance depends on more than one's adequate conceptual training of the immediate object, or to assert a phrase from earlier, it is more than a *reminder*. While both performances can be read as deploying the object O's physical properties as features relevant to the act of persuasion, the latter class has utilized these properties for representational purposes. This is best seen in the difference between pointing to a thermometer as evidence for the claim 'The mercury is up to the 30-line' versus as evidence for 'It is 30F outside'.

Through my position on artifactual reasons, I endorse Brandom's introduction of reliabilism into pragmatism as an important step to a rehabilitation of pragmatist objectivity. However, I depart from Brandom in two essential ways. First, Brandom sees our interactions with the non-linguistic world as "language entry transitions" and "language exit transitions" where the entries and exits are perceptual inputs into and actional outputs from a linguistic-inferential structure (Brandom 1994, 222). In contrast, I see the very *media* or *vehicles* of this inferential structure as the objects of the world, inclusive of marks and noises as well as slaps and kisses.[63] Second, Brandom sees our commitments and entitlements in terms of propositional attitudes like beliefs and their corresponding sentential claims, while I think commitments and entitlements

[62] Levine (2019) gives a concise summary of this conclusion: "Brandom's reliabilism is a *social* externalism rather than an externalism *simpliciter* because reliability is not, he thinks, a concept that can stand independently of scorekeepers *taking* others to be reliable" (71).

[63] I am thinking in the background of Dewey's challenge to Russell's abstractionism: "Is inference the sole mode of distinctive behavior which is not an art, and which does not have its own existential organs, agencies and tools?" (MW 10: 92–93). As I read Dewey, he endorses an inferentialism in which all objects of experience participate.

about the norms often extend outward to the broader variety of meaningful vehicles.[64] Taking these two differences together, when I say that reliable responses of instruments can qualify *as reasons*, I am not merely saying that they are a form of perceptual knowledge, part of a language-entry into the inferential structure. Nor do I see thermometers as giving what amounts to a kind of quasi-assertion. I see thermometers as units of reasoning, as deployable *evidence,* as material contents and physical dispositions embedded within a practice of giving and asking for reasons.

Anticipating this view, Brandom preemptively counters that non-sentential units cannot enter the space of reasons because they lack a standardized meaning. He notes that the propositional form of conceptual content is that which can be negated or conditionalized and goes on to exclude pictures (including presumably the contents of photographs) from this class. He writes, "The distinctive sort of definiteness and determinateness characteristic of this sort of conceptual content becomes vivid when it is contrasted with contents that cannot appear in such sentential compounds, such as that expressed by pictures" (2009a, 220-221). He then goes on to tell a story about why the contents of pictures cannot be conceptual:

> My young son once complained about a park sign consisting of the silhouette of what looked like a Scottish terrier, surrounded by a red circle, with a slash through it. Familiar with the force of prohibition associated with signs of this general form, he wanted to know: "Does this mean 'No Scotties allowed'? Or 'No dogs allowed'? Or 'No animals allowed'? Or 'No pets allowed'?" Indeed. With pictures one has no way of indicating the degree of generality intended. (2009a, 220-221)

Brandom is correct that pictures lack a certain kind of specificity. But so do some sentences. If the sign had read 'No pets allowed', we might run into similar questions. Does picking up a turtle stomping along a footpath and claiming it as one's own constitute a violation of the rule? If not, then can one bring one's pet turtle after all? Or if the sign read 'No swimming' one could perhaps claim that walking into the park's pond up to one's neck without moving one's arms would not technically constitute a violation. What Brandom is giving are questions about rule-following that could only already pertain to those types of entities he treats as conceptual.

Therefore, I think Brandom demonstrates, contrary to his intention, that the picture *is* involved with rule-governed normative practices. The Brandoms understand that 'a red circle with a slash' translates to a prohibition and that the silhouette of the terrier produces a range of possible

[64] This is reflected in Koopman's (2009) claim that pragmatist reliabilism fails when it imposes a "restriction of epistemic honorifics such as justification and knowledge to just those classes of entities resembling propositions, beliefs, and sentences" (123).

interpretations. The circle-slash and silhouette are therefore precisely those contents which can be negated or conditionalized, for we might imagine the same picture without the circle-slash or with a different silhouette. This ability to imagine alternate pictorial possibilities is one which can be publicly exercised by the Brandoms but not by a turtle or terrier (assuming they are not mongerers of pictorial concepts).

Next, I think we can draw a second important lesson here—that these pictorial contents are normative in a specific manner that cannot be translated into natural language. For example, one could know all propositional facts about terriers, pets, animals, parks, red, and circles without being trained to see the sign-designs of the 'red circle-slash' and silhouette as having a particular prohibitive force. This is to say that the Brandoms see the picture as a range of potential assertions because they have been trained to associate marks and noises with and against pictorial concepts. Or alternately, we can say that the Brandoms understand the picture differently from members of other societies who are trained to understand a different range of prohibitory signs.

Finally, the Brandoms are creatures with just the kind of capacities enabling them to deploy the picture as a reason. If a policeman were to stop them walking their German Shepard along the park path, they could point to the ambiguous sign as a defense while perhaps saying (with tonal emphasis on the italics), "That is not a picture of *a dog*", implying that they have interpreted the picture as prohibiting the smaller class of terriers. While Brandom is correct that these actions require knowledge of a natural language, the entire performance is not constituted by language. The move could not be performed by someone who knew English but lacked knowledge of the pictorial conventions of signs (or of the inferential role produced by different tonal emphases within a sentence).

These three lessons—that pictorial contents are conceptual, that they require separate training which is not necessarily reducible to the marks and noises of natural languages, and that they can deployed as components of reasons—apply particularly well to photographs.

### *3.4.2 – Photographic Evidence*

I now return to Ayers' example and imagine two interlocutors arguing over the JFK assassination. In an argument, A attempts to persuade B that JFK was assassinated by showing B the Zapruder film O. In considering how this performance figures into the space of reasons, I want to consider the specific kinds of capacities that are being exercised in A's ability to deploy the Zapruder film

as evidence to B in order to highlight that the complexity of the case cannot be glossed over by perceptual and linguistic reductionism.

First, there is the possibility of *triangulation*. The two interlocutors must be able to triangulate the contents of the film. This is a process that need not involve two language users. It could be performed by any two complex organisms. However, without language or some other kind of (minimally) representational system, there is an inability to deploy the triangulated contents as reasons in an evidential sense. For example, two dogs A and B may both bark at a strange man on the television screen without understanding that the man is neither in the room nor a threat to them. Even if A were to realize that the triangulated contents are not an immediately present object, there is no pattern of barks, wags, or bites that A could perform which would count as an action-guiding reason for reorienting B's behavior with respect to the strange man. In other words, A might be able to intentionally change B's behavior by barking or biting B into silent submission, but these are not surefire ways of indicating to B that the strange man is not in the room—these actions are instances of behavioral modifications of force rather than of persuasion.

With the introduction of language, this first problem is solved. We can imagine that interlocutors A and B understand how to speak about the contents of the film and therefore can make claims about the film, claims which can be linked back to particular pointings at particular timeframes. And so, in a second stage, the introduction of a sufficiently rich system of symbols or representations—in our case a language—allows for the contents of the film to be deployed *as evidence*.[65] But still, this is not enough. We can imagine that, at this stage, A and B can only talk about the *contents* of the film but do not possess concepts pertaining to the reliability of the medium of the film itself. And so, A points to a particular frame of the film as justification for the claim that 'JFK is dead' only for B to point to a hand drawn picture of an alive-and-well JFK as counter evidence.

What A and B both need to understand is that the film is a *reliable vehicle* of its presented contents. In this manner, when A points to a particular frame of the film as justification for the claim 'JFK is dead', B cannot adequately respond by pointing to a hand-drawn picture but must point to either a different frame of the film or to different contents from an equally reliable source.

[65] Recall from the last chapter, that my stipulation of 'representational' is in the deflationary sense in which one object can stand in for another, or from the standpoint of a practice, when the rules of a practice are modified, not through conduct and counter-conduct, but due to modes of conduct which can directly express and modify *the rules* of such conduct. This is my distinction between the space of practice and the space of reasons.

But even here, we do not have a rich enough description of what might happen. Imagine B were to claim that 'the Zapruder film shows a *1965* Lincoln Continental convertible', thereby proving that the film was manufactured after JFK's supposed death. Here, if A does not possess the adequate concepts for being able to demarcate the differences between the 1961 and 1965 models of the Lincoln Continental convertible, then A cannot accept the contents of the film as justifying or failing to justify B's corresponding claim. And so, what A and B need is *triangulation*, adequate *representational resources*, agreement on the *reliability of the vehicle of expression*, and a *shared conceptual background*, conceptual not only in shared linguistic meaning but also in terms of how linguistic meanings are associated with other classes of public objects, visual signs, and other artifacts.

This picture envisions a new model of conversational justification in which tossing a photograph into a conversation *justifies* in the literal sense of the term. Through triangulation we can conceive of two interlocutors, not as self-contained doxastic systems with a causal gulf between them, but as embodied individuals, stuck together in shared practices, capable of deploying chunks of the world directly within their justificatory performances. Using McDowellian language, since these chunks are "inside" the space of reasons, the required friction between the space of reasons and the causal world is already accounted for. This is because the space of reasons extends into and relies upon the rich and propertied landscape of the world of common sense, rather than being contained within the bare system of beliefs, sentences, and vocabularies.

This kind of language may make my position seem like an inflated metaphysics rather than a methodological proposal. But on the contrary, the space of reasons is not an inflated, idealist world-mind. It is just a functional metaphor. The space of reasons is the result of reliable dispositions of organic brains and environing features linking and mixing together over millions of years, a process which nonetheless enabled the complex polydimensional discourse of the drawing room. My claim is that this space is fully functional as a result of the physical properties which have enabled it, not in spite of them. To speak of the physical world as "inside" the space of reasons, is just to clarify that we think *using* the objects of the world rather than only ever with strings of sentences. It is just to say that we can, and most certainly do, *use* photographs as reasons.

### 3.5 – Conclusion

I began this chapter with a discussion of Rorty's conversationalism in relation to diverse forms of expression found on social media. Caught up in R.W. Sleeper's (1985) proposal that there must be something outside of the conversation, something *had* through Dewey's navigational instruments, by a notion of experience, I have been suggesting an alternative. By effectively bringing the rest of the world into the space of reasons, I argue that we can diffuse the worry that we need something external to the conversation to ground our beliefs and desires—Neurath's boat has now been retrofitted with planks of steel and lead and titanium.

This leads me to agree with Van Den Eede (2010) that Rorty's conversation is not strictly an "epistemological mechanism" (197). But I think this is precisely the problem. The propositional bias shared by Rorty, Davidson, and Brandom fails to locate how alternate forms of non-sentential evidence can be taken up into our practices of reasoning, and even when this propositional bias has been challenged within pragmatism, the conditions an artifact must meet in order to count as a reason or to be included within practices of justification have not always been clear.

In endorsing an artifactual pragmatism, I have expanded the space of reasons into our other communicative practices, enabling the contents of these non-sentential practices to serve as evidence. Here, my view comes closest to John McDowell's in the scope and application it affords to an expansive conception of rationality. But where I depart from McDowell is on the insistence that the conceptualized and rationalized contents of the space of reasons are the contents of states of perception. My model is sociological rather than psychological. It treats reasoning as a practice fully explainable within the vocabulary immediately available to the anthropologist and the historian, the politician and the brick layer.

With my theory of artifactual pragmatism in full view, I turn now in Part II to multimodal large language models and audio-visual deepfakes as important applications for understanding the stakes of my theory in the real world. In Chapter 4, I challenge the view that multimodal large language models count as participants in the space of reasons by using Davidson's externalism to undermine the thesis that multimodality can provide non-linguistic material for reasoning. Then, in Chapter 5, I consider cases of multimodal social verification of audio-visual media. Both of these projects notably toe the line between the narrow communicative dimensions of linguistic pragmatism and the broader aesthetic dimensions captured by experiential models of objectivity.

Having broadened the horizon of communicative-theoretic pragmatism to allow for a wide range of performances to count as evidence, I now hope to put artifactual pragmatism to the test.

# PART II:

# PRAGMATISM AND SYNTHETIC MEDIA

# CHAPTER 4

## ALIEN REASONERS:

## TRIANGULATION, MULTIMODALITY, AND CONVERSATIONAL AI

> Man. A Feeling, reflecting, thinking being, who freely walks the earth, who seems to be at the head of all other animals whom he dominates, who lives in society, who has invented the sciences and the arts, who has his particular goodness and badness, who has given himself masters, who has made laws for himself, etc.
>
> — Denis Diderot, "Homme" in *Encyclopedie* (1765)

### 4.1 – Introduction

The central insight of the general program laid out in Part I is that the space of practice and the space of reasons are intricately entangled. Modifications to the artifactual arrangements which make up our practical life produce parallel modifications to "players" and "objects at play" within the space of reasons, as when thermometers, photographs, litmus tests, and even new concepts are introduced to stabilize and adjudicate epistemic contestations which might otherwise remain unresolved.[66] There is a continuous feedback loop between causes, practices, and reasons during the development of new kinds of epistemic artifacts. The artificer plays several "games" at once, talking and thinking with their peers while also tinkering and testing out their prototypes, all while the world pushes back in unexpected ways. At the conclusion of this process, when an epistemically relevant artifact is produced, distributed, and recognized as reliable among a community of reasoners, these reasoners subsequently change their behavior in privileged cases of justification. Over several iterations of this process, linguistic testimony is supplemented with all kinds of instruments—thermometers, barometers, litmus tests, photographs, films, and other recording devices. Minor disputes are settled because the "pointing game" can now pick out a special class of mechanical objects which have no epistemic bias, and which cannot take sides.

[66] While my introductory summary focuses on how new "objects" may be introduced into the space of reasons, one could also study how certain concepts or conducts have brought once-excluded persons into the space of reasons as well. See Susan Dieleman's (2025) account on the status of women and the space of reasons for how Brandom lacks an account of past epistemic and rational exclusions. The question of LLMs in this chapter may also be seen as a contestation over a new kind of rational "player."

But this ideal scenario is easily complicated. Old artifacts break down and lose their reliability, and new artifacts may undermine or disrupt the role that the old artifacts once played. Additionally, some artifacts may no longer operate according to a transparent, mechanical process, but through complex and opaque procedures which prevent them from serving as unbiased epistemic arbitrators.

Many digital technologies which utilize machine learning (ML) operate in this opaque manner.[67] In the past decade, search engines and recommendation algorithms have been two of the most widely-discussed cases in academia and the broader public sphere.[68] But with the launch of the generative AI chatbot ChatGPT by OpenAI in 2022 (OpenAI 2022), large language models (LLMs) have gained increasing attention as a new class of epistemic disruptor. What makes LLMs especially unique is that their "conversational" abilities have become increasingly indistinguishable from human actors. Research studies have shown that ChatGPT-generated essays fool 30% of college instructors (Waltzer et al. 2024) and that fake abstracts fool 32% of trained scientists (Gao et al. 2023). More recently, one study found that 54% of human participants believed that the latest GPT-4 model was a human interlocutor compared to 67% rates of identification for actual human interlocutors (Jones and Bergen 2026).[69] These successes have prompted philosophers to reconsider the "player" question in relation to conversational AI.[70] Specifically, one might ask: are current or future LLMs capable of acting as participants in the space of reasons?

This specific question has gained notoriety in subsequent years, bolstered by the popularity and effectiveness of LLMs, with some philosophers endorsing the position that LLMs merely

[67] For some general accounts on the opacity of computational processes, see Durán and Formanek (2018). For applications of the opacity debate to ML, specifically machine vision, see Alvarado (2022).

[68] Safiya Noble (2018) provides the definitive account of racial bias in search engine results. Issues of polarization and personalization resulting from social media recommendation algorithms were popularized in Pariser (2011). Nguyen (2020) helpfully distinguishes between algorithmic as opposed to attitudinal causes of polarization with his distinction of epistemic bubbles from echo chambers.

[69] These studies are a variation of the "epistemic arms race" I address in relation to deepfakes in Chapter 5.

[70] For linguistic variety, I shift between terms such as 'artificial system', 'generative AI', 'AI agent', 'conversational AI', and 'LLM'. For clarity, each of these terms describes an ever-specific kind of technology. *Artificial system* describes any kind of computational technology which could be interpreted or ascribed intentionality. *Generative AI* covers ML-enabled text, image, audio, or video generators. *AI agent* and *conversational AI* refer to ML-enabled generative AI which produce primarily text-based responses. Finally, *LLMs* are a particular class of conversational AI relying on transformer models.

simulate reasoning (Heinrichs and Knell 2021; Vallor 2024; Sambrotta 2025) and others arguing that the capabilities of AI cannot be so readily dismissed (Søgaard 2022; Marchettoni, 2025; Havlík 2025). This debate extends across positions in philosophy and linguistics concerning semantic content, intentionality, and the role of experience in relation to language (see Harnad 1990; Bender and Koller 2020; Mallory 2023). In many of these debates, and all of those directly relevant to the artifactual pragmatism I advanced in Part I, Robert Brandom's conception of sapience and its relation to the space of reasons is a point of contention.[71]

In this chapter, I contribute to this debate through the positions developed in Part I of this project. Despite my criticisms of Brandom's overly linguistic conception of normativity in Chapter 2, I argue that Brandom's ambiguous positions on triangulation are a key component for considering LLMs' inclusion into the space of reasons. Though Brandom modifies his views over time to align closer to the positions on practice provided by the classic pragmatists (see Brandom 2008, 87, 180), his narrow account of reason-giving and discursive intentionality may nevertheless end up granting rational capabilities to LLMs. Throughout this chapter, I draw out the implications of Brandom's views while also providing a way to exclude LLMs from the space of reasons by mobilizing triangulation in relation to the objects we use within the space of reasons. Since LLMs do not play with the same materials as human players, we can effectively question whether their responses count as legitimate reasons.

The structure of the chapter is as follows. In Section 4.2, I outline the recent debates on the status of LLMs in the space of reasons, supplementing and extending each position with relevant material on externalism and intentionality in relation to AI. In Section 4.3, I demonstrate the relevance of triangulation to this debate by tracking Brandom's conflicting positions on the topic. His earlier work suggests that contact with the world only properly occurs when objects and their states of affairs are picked out by language users, while his subsequent work on analytic pragmatism is more compatible with the view that direct perceptual contact with a world of objects is a pre-requisite for reason-giving. In Section 4.4, I defend my position that the space of reasons draws from the various objects of practice and I use this position to respond to the counter that

[71] Not all of the accounts which specifically refer to the space of reasons cite Brandom as their focal point, but part of the debate has shifted to exegetical analysis of his positions. Heinrichs and Knell initially state that they are only interested in applying the concept from Sellars and Brandom rather than excavating their positions (2021, 1570). And Vallor's (2024) analysis rests on McDowell's experiential variation of the space of reasons as given in his *Mind and Word* (McDowell 1994). Other positions refer to Brandom's inferentialism (Havlík 2025) or his pragmatist account of intentions (Mallory 2023)

multimodal LLMs effectively ground textual responses in the external world. I then briefly conclude with suggestions on how pragmatists should analyze the normative status of future artificial agents.

### 4.2 – Player or Object?

In recent years, beginning some fifteen years after the publication of *Making It Explicit*, Robert Brandom has directed his general program toward philosophical issues in artificial intelligence (see 2008; 2019). This move is not unexpected given Brandom's philosophical commitments. In Chapter 2, I outlined how Brandom distinguishes between *practical intentionality* possessed by nonlinguistic organisms and *discursive intentionality* as "the intentionality of states and expressions with propositional (hence conceptually articulated) contents, such as beliefs and intentions" (2019, 82). This difference is generally aligned with the distinction between beings which possess *sentienc*e as awareness and the ability to feel pain and pleasure and *sapience* as the ability to be normatively autonomous by being able to give and ask for reasons (1994, 275).[72] Centrally, Brandom argues that discursive intentionality is demarcated by means of an *inferentialist semantics*, where the content of some linguistic concept is given by its role within a network of propositions, inclusive of their compatibility and incompatibility relations. As mentioned in Section 2.3, Brandom elucidates this theory through the example of a parrot and the meaning of the word 'red'. While one could train a parrot to say, "That's red" in the presence of red things, it does not know that "That's red" is compatible with "That's colored" and incompatible with "That's green" (1994, 89). Beings which can maneuver through the total field of concepts by means of inference possess the ability to give and ask for reasons and engage in deontic scorekeeping with other beings to track their commitments and entitlements given their expressed linguistic behavior (1994, 142). They are also capable of connecting their true assertions or claims to states of affairs, of representing the facts of the world (1994, 632).

Given this background, one can see how Brandom's positions provide a unique avenue for addressing issues in artificial intelligence. First, by separating sentience from sapience, Brandom distinguishes the ability to know a language from the possession of consciousness or awareness. Second, Brandom's inferentialism treats inference as explanatorily prior to reference in the order

[72] Brandom treats sapience and discursive intentionality as interchangeable concepts (see 1994, 8).

of semantic explanation. Pre-linguistic natural creatures give mere vocal performances which do not represent things in the world until there is an order of complexity and diversity between vocal content to allow for sentential expression and inferential articulation between the component parts of sentences. At this point, vocal content becomes verbal content, and the creatures are able to refer to objects and represent states of affairs. The question which I will address below, however, is whether Brandom's position extends to disembodied artificial systems. If his program allows conversational AI to represent the world, then Brandom's semantics is a possible mechanism for attributing intentionality to artificial systems.

*4.2.1 – Applications of the Space of Reasons to LLMs*

In recent years, philosophers have gauged the potential importance of Brandom's work in relation to the potential rationality of artificial intelligence. In the article "Aliens in the Space of Reasons?," Bert Heinrichs and Sebastian Knell (2021) address the question of AI agents participating in practices of reason-giving, initially concluding that AI agents are only sophisticated tools and not genuine participants in the space of reasons. Their basic position comes down to the fact that these agents do not undertake commitments for their claims. For instance, AI agents "seem unable to understand themselves as standing in genuine normative relationships with other entities and their performances" (1577). Heinrichs and Knell concede, however, that they do not assert this point dogmatically, admitting that it could be possible for future AI agents to undertake commitments (1577). This admission is complicated by the overall advancement in AI's conversational abilities since the publication of their article. The authors' main examples of AI agents include recommendation systems, chess engines, and machine vision models driven by Convolutional Neural Networks (1573–1574), none of which capture the complex, emergent inferential capacities of the latest class of AI agent.

Despite these complications, both Shannon Vallor (2024) and Mirco Sambrotta (2025) have defended this initial position in relation to more sophisticated conversational AI, such as LLMs. Vallor holds that LLMs "mirror" human linguistic capabilities in a manner that allows them to "bypass" the space of reasons by generating textual outputs which appear human (2024, 107). Vallor gives a cluster of arguments against the reasoning capacities of generative AI. One issue lies in the fabrications that conversational AI models produce (25), a result of generating outputs from a series of next-word prediction tasks (49–50), such as the auto-regressive generation shown

in Figure 1.[73] A second related issue is their inability to back up or defend their responses in light of questioning (31). A third issue is their general lack of experience, what she terms "our first and unbroken bond with the world" (59). For Vallor, our experience ultimately grounds our thinking and brings our reasons out into the light of day. She states, "We relate to other minds not as mirrored surfaces, but as mutual *experiencers* of a common, open world" (59). This notion of experience plays a central role in the meaningfulness of our linguistic testimony. Ultimately, our status as mutual experiencers is bound up with our ability to inhabit and navigate the space of reasons. Without the capacity to navigate the world writ large, generative AI fundamentally lack the ability to navigate the space of reasons.

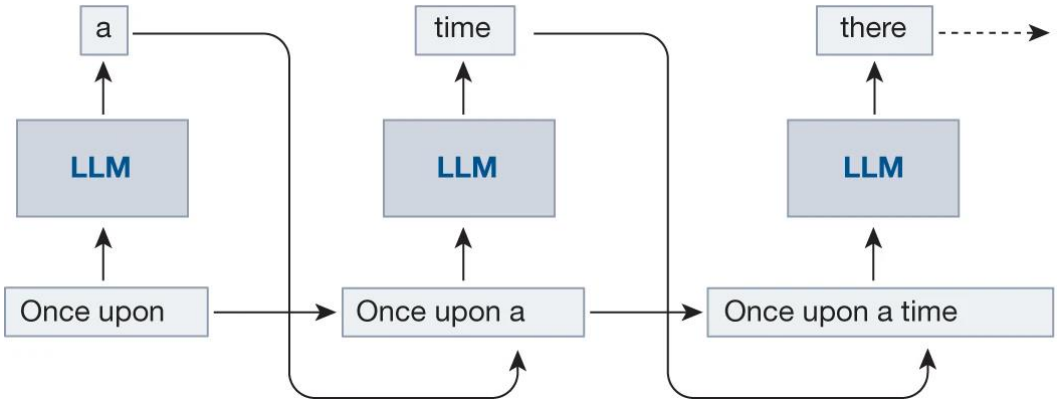


Figure 1: An example of how LLMs generate text through next-word prediction. (Shanahan et al. 2023)

In response to both skeptical positions, Leonardo Marchettoni (2025) argues that neither view has correctly assessed the implications of Brandom's inferentialist semantics, noting that Heinrichs and Knell do not describe agents with sufficient inferential capacities and that Vallor has glossed over such capacities by explaining LLMs' reasoning using lower-order functional explanations. Marchettoni makes two novel moves against these positions. First, he argues that comparisons of human linguistic abilities to the underlying statistical processes of generative AI are a kind of category error akin to comparing the linguistic capacities of LLMs to patterns of human neural activity (2025, 7). For the human-machine comparison to adequately address the question at hand, one would need to restrict oneself to descriptions of only the overtly interpretable

[73] A comprehensive introduction to ChatGPT is provided by Steven Wolfram (2023).

linguistic behavior in line with the model of the Turing Test. Second, Marchettoni thinks that analyses which explain away LLMs' ability to give reasons by referring to external causal explanations make a similar mistake. He asserts that the capacity to give reasons are to be assessed independently of any adequate grounding condition, that "what really matters is what inferential moves a given agent is capable of making, and whether its capabilities enable it to play the game of giving and asking for reasons" (10). In this manner, Marchettoni thinks that an LLM's capacity to provide true statements to a wide variety of interlocutors and to provide further statements upon questioning minimally counts as genuine participation in the space of reasons despite the unusual causal processes underlying these responses.

On the other hand, Mirco Sambrotta (2025) claims that while LLMs may make adequate inferences at the surface level, they fail at the level of scorekeeping. The issue with a purely behavioral analysis is that it misses Brandom's conception of normative autonomy. Since LLMs only provide justifications for their statements by drawing from their training data, they are not actually responding to their interlocutors, nor updating an internalized set of commitments and entitlements (2025, 13). If Sambrotta is correct, then Marchettoni misses that the ability for a system to make correct inferences is necessary but not sufficient to ascriptions of discursive intentionality.

But the skeptical position requires a further elucidation on the epistemic status of LLMs. An important feature of these conversational AI models is how their capacity to give truth-evaluable textual outputs have made them a new class of epistemic artifact or instrument. In the past, thermometers were a typical example of an instrument socially sanctioned or indexed to respond to a particular class of stimuli (see Dretske 1989, 10–11; Brandom, 1994, Ch. 4; Chang 2007; Goldberg 2020), but thermometers could not produce a variety of linguistic outputs, nor could they tailor their responses to the particularized inputs of their human interlocutors. Even some of the most critical positions on LLMs assert that they fool humans by generating text-based outputs that appear to be truth-evaluable *propositions* (Fisher 2024, 4; Hicks et al. 2024, n6). This renders them an unusually flexible kind of epistemic artifact. If LLMs lack discursive intentionality, then how should their novelty be conceptualized relative to other kinds of objects at play in the space of reasons?

Brandom himself argues that "computers are already second–class members of our discursive community" in the sense that they extend perception and report phenomena to us (in

Frápoli and Wischen 2019, 651). Humans have already been "anthropomorphizing" computational chatbots since at least Joseph Weizenbaum's (1966) decision-tree-based therapy chatbot ELIZA. New kinds of AI agent would simply be an extension of this phenomenon. In the contemporary literature, artificial agents have been treated as props within games of "make believe" (Mallory 2023) or are said to function like improvisational actors, using the processes of auto-regressive generation to continuously string a coherent performance together (Shanahan et al. 2023).

Within this general paradigm, the comparative novelty of LLMs comes from their capacity to produce linguistic content which is truth-evaluable, content which sits between the limited class of properties to which technical instruments respond and full-blown human testimony. Various concepts have been developed to chart out this space, including "quasi-testimony" (Freiman and Miller 2020), "technology-based beliefs" (Freiman 2023), and "proxy-assertion" (Arora 2024).[74] These positions are readily available to the Sellarsian pragmatist interested in the question of LLM reasoners, and it seems like Brandom himself is willing to endorse such an account.[75] Yet the problem remains as to whether these positions can be maintained in light of the broader debates emerging around the increasingly human-like intentionality of conversational AI.

*4.2.2 – Externalism, Intentionalism, and Multimodal LLMs*

The wider framework regarding LLM intentionality and meaning problematizes the status of artificial reasoners. Outlining the general strands of argument, Fintan Mallory (2023) gives two forms of argument which are relevant to the above debate: *semantic externalism*, which addresses how the content of language is derived from particular causal relations to the world, and *intentionalism*, which addresses the underlying mental states an entity must possess to assert something.[76] Mapping the above positions to each form of argument, Vallor's (2024) position on experiential grounding is a variant of an externalist argument while the Marchettoni (2025) and Sambrotta (2025) debate centers around ascriptions of intentionality to LLMs.

[74] Abbas Bagwala (2026) similarly argues from a Brandomian perspective that LLMs possess *proxy sapience.*

[75] However, to note one account that clashes with Brandom's sapience-sentience distinction, Williams and Bayne (2024) argue that LLMs are "proto-asserters" which possess some degree of discursive intentionality. This position is more difficult to reconcile with Brandom and Sellars' commitments.

[76] Semantic externalism is often intertwined with *content externalism*, the view that mental content is produced by factors external to the mind or body (see Ebbs 2025).

Positions on semantic externalism come in several varieties. In philosophy, the position is canonically associated with the work of Hilary Putnam (1975) and Saul Kripke (1980). Yet one of the most pressing externalist criticisms in the contemporary debate comes from Emily Bender and Alexander Koller (2020), who rely on Steven Harnad's (1990) conception of the *symbol grounding problem*. In Harnad's original article, the problem is intended to supplement John Searle's classic "Chinese Room" (Searle 1980). Though Searle shows that cognition cannot be merely the use of symbols according to rules, Harnad is interested in showing how the problem is connected to a lack of perceptual grounding. Linguistic symbols must be grounded in *iconic representations* or "analogs of the proximal sensory projections of distal objects and events" (335) as well as *categorical representations*, which allow us to identify objects as pertaining to a given category (342).[77]

Bender and Koller (2020) modify this basic notion of symbolic grounding to argue that sub-symbolic generative AI still lacks meaning. In what they call "the Octopus Test," a hyper-intelligent octopus eavesdrops on A and B's conversation through an undersea cable before attempting to impersonate B. Bender and Koller argue that the Octopus may effectively mimic the form of the language from learning past cases, but that it lacks the meaning since it has never interacted with the referents of the speech (2020, 5188). Given the Octopus' inability to learn the meaning of the language, Bender and Koller assert that "the process of acquiring a linguistic system, like human communication generally, relies on joint attention and intersubjectivity: the ability to be aware of what another human is attending to and guess what they are intending to communicate" (5190). Since the Octopus cannot access the intersubjectively available objects of A and B's social world, it therefore merely mimics a language. On this model, conversational AI models are, to use Bender's famous phrase, "stochastic parrots" that merely mimic language by repeating the statistical patterns of their pre-training datasets (Bender et al. 2021).

However, the Octopus Test is a source of controversy within debates on whether LLMs know the meaning of a language. The general criticism is that Bender and Koller have confused possession of awareness with linguistic meaning, that knowing how to use words relative to other words is sufficient to know a language even if that entity does not possess something like "qualia"

[77] Harnad distinguishes between perceptual discrimination and *identification* where the latter involves come kind of conceptual processing of iconic perceptual intake into classes (1990, 341). In many respects, Harnad's notion of iconic representation is heavily Quinean and his notion of categorical representations is reminiscent of McDowell's (2008) later conceptual (but non-propositional) account of perception.

(Gubelmann 2023, 487–488; Søgaard 2023, 40). Additionally, Anders Søgaard argues that LLMs can be grounded on language alone due to evidence that their representations are isomorphic to the physical world (2023, 43). Similarly, Vladimir Havlík, drawing from Brandom, endorses a position clearly within an inferential lineage: "If an expression is to represent something, then it represents its place in the complexity of relations to all other expressions" (2025, 17). Piantadosi and Hill (2022) extend this account even further, holding the position that a conceptual role semantics (of which inferentialism is a kind) can provide LLMs with inner mental states, especially when textual data is supplemented with multimodal training data. This proposal may even imply that iconic representations provided by the multimodal data themselves become material possessing a conceptual role when these data are properly integrated into the pre-existing representational space of an LLM.[78]

Indeed, a counter to the symbol grounding problem is that LLMs are no longer merely text-based artificial systems and therefore, that the sphere of conceptual has shifted outward to encompass new kinds of content. Most LLMs are now supplemented with Vision Transformers (ViT) to read and produce image-based content and some even possess Retrieval-Augmented Generation (RAG) to bracket out textual descriptions from specified source material.[79] Multimodal LLMs (MLLMs) may meet the challenge posed by externalist theories of semantic content by "grounding" textual relationships within audio, video, and image data (see Figure 2). For instance, Williams and Bayne (2024) claim that "objections from the text-bound nature of chatbots evaporate in the face of the fact that most of the leading systems are now 'multimodal' – processing text, audio and video inputs" (2024, 14). RAG-enabled systems may similarly solve issues of normative autonomy and commitment by giving LLMs a particular standpoint generated from uploaded source inputs. At least one strong skeptic of LLMs, Sarah Fisher (2023), has argued that the problems posed by auto-regressive text generation are solved when "a proposition is retrieved in this way and served up to the user" (2024, 6).

[78] Though I want to reject the conflation of conceptual understanding with sentience, it should be noted that David Chalmers believes that MLLMs are a possible route to conscious AI: "Because human consciousness is multimodal and is deeply bound up with action, it is arguable that these extended systems are more promising than pure LLMs as candidates for humanlike consciousness" (Chalmers 2023).

[79] ChatGPT, Claude, and Gemini are all MLLMs. NotebookLM is an example of a RAG-enabled MLLM. For a general background on Retrieval Augmented Generation see (Gao et al. 2024; Huang et al., 2024).

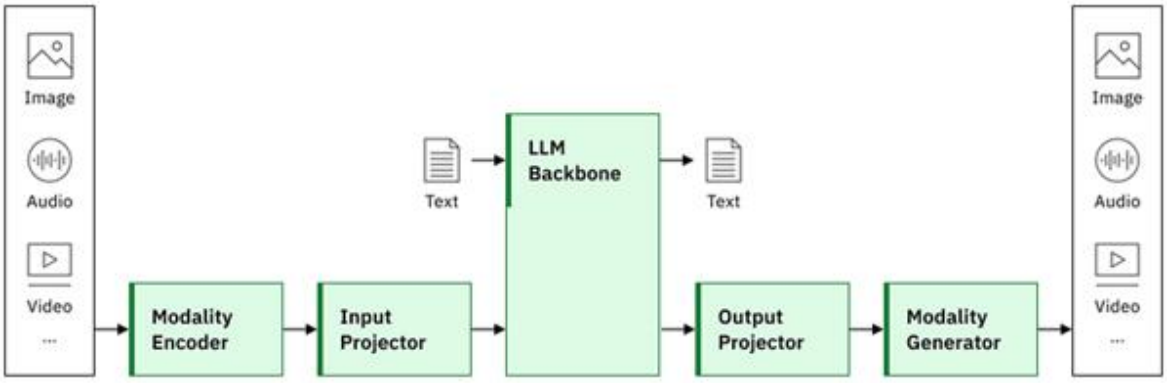


Figure 2: An LLM supplemented with multimodal capabilities. (Varughese 2025)

These modifications are but the start of a long line of potential supplementations to base LLM architectures. The question is how the externalist and intentionalist critics of LLMs as reason-givers should respond to these updates. For Vallor (2024), the problem is that her account of the space of reasons is dependent upon a notion of experience and therefore, suffers from the criticisms posed against Bender and Koller. For Sambrotta (2025), the issue is whether supplementation of LLMs with novel source inputs using RAG is enough to generate basic commitments beyond the initial patterns generated during pre-training. These challenges demonstrate that LLMs may not be as readily excluded from the space of reasons as many have thought.

### 4.3 – Triangulation and Analytic Pragmatism

Given the context above, I turn now to a partial vindication of Vallor's (2024) position. I expand upon the space of reasons from the externalist angle, one which would require that any future artificial system be embodied in the world as a pre-condition for entry into the space of reasons. To this end, I draw from a very different part of Brandom's theoretical edifice—his utilization of the conception of triangulation. Though Brandom has not been read as an externalist (see Mallory 2023, 1084–1085) and has even been described by some as an opponent to externalist arguments (see Marchettoni 2025; Havlík 2025), Brandom's positions are ambiguous as a result of his shifting views on triangulation within his developing program on AI (2008, 2019).

As discussed in the previous chapter, the initial position of triangulation was developed by Donald Davidson as a means to rescue objectivity. Against Quine's position that organisms only respond to the proximal stimulus of their own sensory organs, Davidson (1990b) holds that social organisms responding to each other, to a common object, and to each other's responses *triangulate* a common object of the perception. Davidson's position, though not directly cited, provides the general style of argument taken up by Bender and Koller (2021) in their Octopus Test and then later by Vallor (2024). Importantly, Davidson also does not conflate awareness with linguistic understanding, allowing his model of triangulation to be a position which causally grounds symbols in the world without pre-supposing that sapience requires sentience.

However, since Davidson does not formally use the Sellarsian notion of the space of reasons, it is difficult to see how this notion may be tied to the game of giving and asking for reasons. Brandom does utilize this concept, though with some ambiguous results for the philosophy of AI. In this section, I outline how Davidson uses triangulation in his later work to develop a variation of the Turing Test which resembles Bender and Koller's Octopus Test. Then, I turn to how Brandom takes up this notion relative to his explicit positions on AI.

### *4.3.1 – Davidson's Turing Test*

Davidson's formulation of triangulation is essential for understanding how he repositions the body of organisms and the contents of their shared environment as requisite features for attributing meaning to their utterances. After the initial development of these themes, Davidson explicitly applies them to considerations of artificial intelligence. Though Alan Turing (1950) had devised a test for intelligence based entirely upon conversational capacities of an artificial system, Davidson utilizes triangulation to show that the scope of the Turing Test is too narrow (1990c). Conversational abilities cannot alone determine whether the history of some interpretable system is sufficiently hooked up with the world to provide meaningful responses. Davidson points out that Turing's test already presupposes that the system "be causally responsible for the 'messages' observed by the interrogator" (1990c, 83). But he holds that an additional consideration of the system's body must be included to see how it reacts to the world (1990c, 83). Leveraging triangulation, Davidson writes that "it must operate in the world in a certain way, and for someone else to grasp those semantics, there must be a three-way interaction among object, interrogator, and a shared world" (1990c, 83–84). Davidson's understanding of the body presupposes an active

being which is constantly interfacing with shared objects. Not only must an artificial system "be brought into the open so that its causal connections with the rest of the world as well as with the interrogator can be observed," but in keeping with the standards of the Turing Test, it "must be made indistinguishable from a person from the interrogator's point of view" (1990c, 84).

For Davidson, the triangular relation also sets up three co-constitutive kinds of knowledge: knowledge of oneself (subjectivity), of others (intersubjectivity), and of the world (objectivity) (1998, 87). Within the triangle, a creature understands its own position in the world only relative to the presence of another interpretable creature. At the same time, in order to identify this fellow creature, the first creature must distinguish the fellow creature's body from the objects of their shared environment. They must also understand their fellow creature to be undertaking the same interpretive stance toward them. In this manner, triangulation treats the first-person, second-person, and third-person points of view as mutually dependent; a person only understands the content of their own thoughts by understanding their position in a world which includes other persons and objects.

The result of Davidson's position on the Turing Test is fairly radical, for it would treat any entity that cannot replicate a broad range of human linguistic and non-linguistic behavior as failing to possess intrinsic meaning. The interpretable system would need to know a language in addition to possessing human-like movements, gestures, and expressions. It would also have to display behaviors which suggest that it is equally engaged in practices of interpretation. Perhaps to prefigure his example in relation to Turing's Test, Davidson presents commitments which are too strong for general purposes. An interpretable system need not be *indistinguishable* from a person in the sense that it possesses a human body. Rather, given his overall program, Davidson's main point is more constrained: that no system which is disembodied or detached from the standard environmental conditions of the human language-learner could possess the same linguistic content as its human interpreter. Davidson's positions on interpretation are therefore intertwined with his position on the external content of language—an artificial system cannot know a language without knowing its way around the world in general.[80]

[80] Davidson's position on the Turing Test is connected to his famous Swampman example, where a lightning strike decomposes the original Davidson and reproduces, atom-by-atom, a replica of him (Davidson 1987, 19). Davidson's controversial point in this thought experiment is that his doppelgänger would not possess adequately organized thoughts since it would lack the causal history to the world that the original Davidson possessed. Surely, this example is radical, but one might see the insight in terms of a new clock that has been inadvertently wound to the right time.

*4.3.2 – Brandom on Triangulation*

Brandom's uptake of Davidson's positions are ambiguous. On one hand, Brandom openly endorses the Turing Test as a measure of linguistic meaning in a manner that separates linguistic behavior from non-linguistic behavior and treats linguistic representation as the primary mode of access to the world. On the other hand, Brandom claims that his positions require a pre-linguistic practical situatedness in the world which intermixes discursive practices with their lumpy referents. Brandom develops the first position in opposition to Davidson's conception of triangulation while Brandom's latter position incorporates triangulation as a steppingstone toward discursive intentionality. I argue that Brandom's position on triangulation is the crucial criterion as to whether his position would consider LLMs to be participants in the space of reasons, but that, it remains unclear which part of his conflicting positions takes precedence over the other.

Regarding his more skeptical uptake of triangulation, Brandom (2008) openly endorses a variant of the Turing Test for assessing whether AI possesses discursive intentionality or sapience. He states that:

> There just is no point in insisting that something that is genuinely indistinguishable (including, crucially, dispositionally counterfactually) from other discursive practitioners in conversation—no matter how extended and wide-ranging in topic—should nonetheless not be counted as *really* talking, so thinking (out loud), and deploying a meaningful vocabulary. (Brandom 2008, 74)

He later qualifies his position more explicitly: "there is no reason why electronic computers could not be discursive creatures" (quoted in Frápoli and Wischin 2019, 650). These statements suggest that Brandom, in contrast to Davidson, is willing to separate the rules implicit within linguistic behavior from the total exhibited behavior of some interpretable object. While Davidson's modified test fundamentally undermines the ability to distinguish "discursive practitioners" and their "conversation" from their total set of interactions in the world, Brandom has no issue carving up behavior along discursive versus practical lines.

The disagreement is more pronounced where Brandom reconstructs what he terms Davidson's *causal triangulation* as a competitor to his own *conceptual triangulation* or *inferential triangulation*. Though Davidson is perfectly comfortable treating triangulation as a pre-cursor to language use, and so treats many complex social animals as having this capacity, Brandom thinks

---

Though the immediately observable functions of the clock look right, knowing that it was never intentionally indexed to the time would show how it effectively "simulates" correct time-telling.

that this kind of pre-discursive causal triangulation underdetermines the common object (1994, 430). The only way for two creatures to refer to a common object in the world is to possess relevant concepts which are inferentially articulated. In this manner, the only way one could tell that two creatures triangulate a tree as opposed to a plant would be that such creatures might say things such as "Look at its big trunk!" and "Is it an oak?" where both linguistic performances can only make sense with 'its' and 'it' substituted by 'tree'. Without these kinds of performances, Brandom thinks that causal triangulation cannot perform the adequate work of directing social organisms to common objects.

However, Brandom's later commitments complicate his distance from Davidson. His disagreements with Turing on the status of AI emphasize his underlying pragmatist commitments. Classic functionalist accounts of AI treat discursive practices as algorithmically decomposable into non-discursive practices in the same manner as long division can be broken down into practices of multiplication and subtraction (2008, 76). In contrast to this *algorithmic elaboration*, "where exercising the target ability just is exercising the right basic abilities in the right order and under the right circumstances" (2008, 26), Brandom argues that discursive practices emerge from a "*pedagogical* practical elaboration" which involves a process of training (86). For this position, he draws from his Hegelian and pragmatist background: "I think of training as a course of experience, in Hegel's and Dewey's sense (processual, developmental Erfahrung rather than episodic, self-intimating Erlebnis) of a feedback loop of perception, responsive performance, and perception of the results of the performance" (2008, 87).

This appeal to the feedback loop of practical activity leads Brandom to endorse a more gradual shift from causal triangulation to conceptual triangulation. He explains that "representational purport" within an inferentially-articulated discursive practice "involves a kind of triangulation on the object represented that is evidently the discursive, theoretical culmination of a progression that begins with the way physical objects are incorporated in feedback-governed practical engagements with things" (2008, 188). This means that Brandom, despite his sharp distinction between sentience and sapience, may be equally committed to a position on the necessity of embodiment for discursive intentionality as are Davidson and even Dewey.

The transition from the initial causal interaction with the world to full-blooded discursive practices begins with a feedback loop of causal, practical triangulations. Brandom asserts:

> The most basic sort of practical triangulation on objects happens when the result of one responsive performance is itself the stimulus eliciting a further response—as, for instance, with the nail one hits with the hammer, perceives the effect on, and then responds to anew by hitting it again, until it is driven flush. (188)

He continues that the next step involves differential commitments of "material inference-and-incompatibility," as when one associates the use of the word 'acid' with sour taste and the color red produced by phenolphthalein (188). The move from the first position to the middle position is exemplified by continuous application of what Brandom calls the TOTE or Test-Operate-Test-Exit cycle, where feedback from one practical action is used to update future action in relation to a set goal (178).[81] Brandom treats continuous loops of this TOTE cycle to be adequately world-involving, such that "feedback-governed practices are 'thick', in the sense of essentially involving objects, events, and worldly states of affairs. Bits of the world are incorporated in such practices, in the exercise of such abilities" (178).[82] At the end of this process of continuous practical updating to the environment, after vocal performances are intermixed into the feedback loops, practices consist of "purely inferential-and-incompatibility relations" between linguistic concepts (188). This is where Brandom thinks that *representations* of states of affairs comes into the picture, where immediate practical and perceptual limitations are transcended and claims can now be made about distant and unobservable states of affairs. Yet at the point where discursive intentionality is achieved, Brandom still wants these practices to be embedded within the prior cycles of feedback relationships. Regarding the linguistic concepts for 'hammer' and 'screw', he states that "you cannot say what hammering and screwing are without referring to the actual objects incorporated in them in different ways: the hammers, nails, and so on, that play essential roles in those practices" (179).[83]

Brandom's practical elaboration of discursive intentionality in relation to AI complicates the philosophical stances often attributed to him. On one hand, in Chapter 2, I showed how

[81] Brandom's model for the TOTE cycle is Dewey's early criticism of the concept of the "reflex arc" (see Brandom 2008, 180; RAC, EW 5: 96–109).

[82] Compare this statement to Brandom's similar account at the conclusion of *Making It Explicit* (1994, 632).

[83] Here, Brandom's analysis of tools looks more Deweyan. For instance, Dewey writes, "Man's bias towards himself easily leads him to think of a tool solely in relation to himself, to his hand and eyes, but its primary relationship is toward other external things, as the hammer to the nail, and the plow to the soil. Only through this objective bond does it sustain relation to man himself and his activities. A tool denotes a perception and acknowledgment of sequential bonds in nature" (EN, LW 1: 101).

Brandom's demarcation of sapience and sentience failed to articulate the inherent normativity in our use of artifacts. Even in the above passages Brandom treats artifactual performances as a stepping-stone to fully discursive intentionality, and his development of implicit norms seems dependent upon their capacity to be made explicit using linguistic concepts.[84] Yet Brandom's reformulation of his position on triangulation from a pragmatic starting point shows that he may not be dispensing with Davidson's suggestions as much as he is updating them to ensure that triangulation can play an adequate role in a full-scale semantic theory. If this is the case, then there is a continuity between the original, unfocused processes of causal triangulation and the representational and objective purport of conceptual triangulation.

### 4.4 – Triangulating Reasons

The above analysis shows that Brandom has tempered his positions against triangulation and is more willing to gradually move from causal triangulation to conceptual triangulation. Yet from the perspective of my artifactual pragmatism, I worry that, so long as Brandom's understanding of conceptual contentfulness only involves the use of words or symbols, and not other non-word triangulated objects, his conception of the space of reasons can become too readily detached from prior perceptual and world-involving relations. If this is the case, Brandom's account of causal triangulation merely requires that some artificial discursive system initially be causally connected to its objects. But this requirement would not bar LLMs from participation in the space of reasons, since it could pick out states of affairs through higher-order conceptual triangulation these states of affairs through the causal processes of its physical realization in hardware.

This is to say that though LLMs are related to parts of the world in an opaque and indirect manner, they are still causally situated in the physical world (for instance, within a distributed cluster of data centers). Regarding practical elaboration of conceptual norms, LLMs may be said to undergo Test-Operate-Test-Exit processes during their pre-training in a way that would ensure that their discursive intentionality arises from non-algorithmic processes. This practical elaboration is further demonstrated by LLMs' base sub-symbolic neural architectures which are radical different in scope from earlier symbolic AI relying upon pre-determined rules or lengthy

[84] Though elsewhere Brandom is more careful to separate *practical intentionality* from what he calls *social pragmatism about norms* (Brandom 2019, 84).

decision trees.[85] Finally, the pre-training data and search capabilities of LLMs connect them to parts of the external world even if these are fed to them in tokenized chunks of data rather than in terms of retinal stimulations, cochlear hair vibrations, or firing nociceptors. Just as a blind person might know that "The sky is blue" without ever seeing the color blue, an LLM may equally know linguistic facts without possessing any direct perceptual contact with things. And MLLMs would be like entities that have never seen the sky but have seen images or videos of it. If this view holds, one reading of Brandom would suggest that LLMs, and especially MLLMs, whose outputs produce human-level rates of correct inference between terms, would have genuine contact with true states of affairs regardless of the underlying physical and technical processes through which these inferential relations were realized. Brandom would essentially agree with proponents of LLM intentionality who argue that perceptual contact and embodiment are not necessary.

This potential problem needs to be bypassed by reconsidering the role of triangulation within communicative-theoretic pragmatist accounts. Artifactual pragmatism treats the objects of perceptual triangulation as active parts of the game of giving and asking for reasons rather than just as part of the story of how linguistic intentionality emerges.

One way of linking artifactual pragmatism to Brandom's analysis of triangulation is to suggest that the development of material inference-and-incompatibility relations between objects already yields genuine norms-of-use. This position follows from the previous chapter, where I challenged Brandom's views on the conceptual contentfulness of a park sign by disentangling the notion of contentfulness from that of vagueness. I think a similar issue is at play in relation to his understanding of triangulation as a component part of his pragmatism. Just because a pre-linguistic child cannot say things such as "A cube has six square faces" and "Squares have four sides" does not mean that she cannot causally triangulate the distal stimulus of the red cube. She demonstrates practical mastery and some degree of conceptual mastery over the object when she places the red cube in the square hole of her toybox. Or to use one of my earlier examples, when one effectively uses a Phillips head and a flathead screwdriver in relation to their corresponding screws, one understands the objects related within these practices as conceptually articulated by some set of socially identified relations of functional compatibility and incompatibility.

[85] My views here as to whether LLMs are reducible to algorithmic processes have benefited from lengthy discussions with Abbas Bagala. See Bagwala's (2026) manuscript for a similar discussion on Brandom's account of practical versus algorithmic elaboration of norms.

Inspired by the conceptualist positions of John McDowell (1994), I take it that if the objects at this level of analysis possess a pre-linguistic kind of conceptual contentfulness, then they can justify propositional claims without being epistemic givens. As I read it, concepts are not created when inferential relations are established between linguistic items. Rather, objects are already materially differentiated and classified as part of social practices. Inferential articulation of linguistic concepts modifies the affordances of prior conceptually-structured objects, so that they may function as reasons within the higher-order practice of giving and asking for reasons.

Reframing my position in these terms helps elucidate the different functions triangulation plays within my account in contrast to those of Davidson and Brandom. In my view, triangulation is the mechanism through which perceptual objects are identified for use within practices (the Davidsonian causal insight) and the mechanism through which chains of belief latch onto the world as justification (the Brandomian conceptual insight). Together, these two insights reconfigure triangulation. The concept no longer serves as an anti-skeptical position regarding the relation of language and world. Now, it is a means through which various environmental objects can be introduced and utilized as parts of justificatory performances. Therefore, at the most general level, my move is to redeploy triangulation to shift away from questions concerning the inherent meaning or intentionality of an LLM's outputs and instead consider questions concerning the available resources an LLM uses as reasons.

This move strategically concedes to proponents of LLM intentionality. Externalism may fail as a mechanism for attributing linguistic intentionality or understanding of a language to an LLM, but the space of reasons extends further than language. It draws from all other background practices. Some of our very basic notions of trust and expertise are predicated on differential distributions of non-rational practices and abilities—for instance, the handiness of a bicycle mechanic, the technique of a painter, or the skillset of a welder. These practices should not be regarded as incidental to the space of reasons; they structure persons' capacities to give and receive reasons from each other, and they tailor the uniqueness of their perspectives as reason-givers. When a child learns a language by pointing at and playing with her toys (e.g., a red cube), she will later have recourse to deploy these toys to demonstrate her linguistic knowledge. For example, she will readily use red and cube-like objects to demonstrate her knowledge of the words 'red' and 'cube' to her father. Within these triangular relations, the child understands her own embodied position in the world, the position occupied by her father, and the shared world of objects which

exist independently of them. In possessing these three kinds of knowledge, she is capable of intentionally utilizing external objects to give reasons to her father. Without this practical interplay between a body and other persons and objects, LLMs will be inhibited in their ability to play in the same kind of game.

Environmental and experiential limits to reason-giving occur often. A birder cannot point to a blue jay as a reason for their belief "That's not a blue bird" if their interlocutor is not well adept at distinguishing the two. A pianist cannot prove to their interlocutor that a song is in D by aligning its melody with a D-major chord if their interlocutor does not know the names of the keys on the piano. Those with sensory impairments similarly lack access to some parts of the space of reasons while often supplementing themselves with new kinds of abilities and embodying different kinds of communicative practices like sign language and braille. Every rational being will differentially inhabit the space of reasons and have different available resources for reasoning, from interactions with different kinds of objects to different masteries of language, vocabulary, and signs. Where my skepticism about LLMs as reasoners lies is that their lack of embodiment prevents them from accessing a minimal repertoire of objects utilized in everyday social practices of reason-giving.

Before arguing for this view, I should note that my skepticism extends even to MLLMs, where the issue is that multimodal material serves as an analog for proximal stimulation of sensory organs. It is not triangulated but rather decomposed into tokenized inputs within training. This method does not follow Davidson's position that interpretive interlocutors see each other within a shared environment. Rather, it is closer to a Berkeleyan or Humean empiricism where common objects are decomposed into streams of contiguous sense-data of different modalities and then stitched back together through non-algorithmic, associative mental processes. Just like the atomistic sense-data of old, transformer models break down images into smaller patches (usually 16x16 pixels) during processing before then associating these "tokens" with tokens from other modalities (Dosovitskiy et al. 2021). This process makes machine vision opaque to human interlocutors who see images as holistic Gestalt configurations rather than a processual stream of decomposed information. Operationally, the difference in visual capabilities leads MLLMs to act in strange ways within epistemic tasks. When used for image or video verification, MLLMs search for textual descriptions of the image rather than employing a more human-like "pixel-based

matching" strategy for comparing unknown images against prior known images (Chandrasekar and Jaźwińska, 2025).

Of course, even the capacity to compare images through matching strategies would not account for the lack of embodied engagement with ordinary objects and events. An MLLM which does not participate in object-related practices performs an entirely different language game than a standard human reasoner. While the human could relate words to the tools in a toolbox, an LLM could only relate words to images of tools (and this is already stretching the notion of 'image'). According to artifactual pragmatism, LLMs interact with an entirely different field of artifacts than human beings. And though these discrepancies may not matter to attributions of linguistic intentionality, they do matter to attributions of standard reasoning capabilities to LLMs, especially in cases where they need immediate perceptual experiences of non-linguistic behavior to correctly assess their interlocutors. Here, I agree with Vallor that "gestures and other signs can convey our reasons too" (2024, 107). LLMs are a kind of entity which cannot read bodily gesture, facial expression, and vocal intonation. They cannot point out objects, nod, or look and move around. They cannot directly engage in embodied practices which make up the majority of normative life. As a result, they are an entity whose manner of assessing reasons is compromised.[86] An LLM's reason-giving is compromised in spite of its capacities to produce true propositions. It is compromised *even if* it possesses intentionality, knows a language, or grasps facts about the world.

In Brandom's terms, what I am arguing is that discursive intentionality without practical intentionality is empty; the objects of practical intentionality provide the content through which discursive norms are realized. An artificial system which possessed a minimal linguistic intentionality would be like a child who knew how to play a "kicking game" but never possessed the required ball. Giving the child pictures, sounds, and videos of balls would still not reproduce the norms of the game. Maybe she could speak about the game, give true statements about its norms, elaborate the various kicking techniques, draw pictures of it, but she would not be able to act in accord with rules of the game. The game could never be a legitimate form of life for her.

The fundamental barrier to an LLM's capacity to participate in the space of reasons is that it cannot act according to our practical forms of life. They are inhibited at the level of what I have

[86] This lack of embodiment is especially problematic in relation to assessments of patient emotional and mental health within clinical care settings (see Brown and Halpern 2021, 3). This is not to mention that a lack of sentience may pose an unbridgeable gap for empathetic forms of medical AI (see Montemayor et al. 2022).

called *artifactual normativity*, where our embodied and social engagement with artifacts provides a thick conception of normativity that undergirds linguistic practices. Since LLMs are not directly embedded within a network of interlocking practices, the philosopher's ascriptions of linguistic intentionality to them are surface-level addendums that do not radically alter our conceptions of normative autonomy. Perhaps LLMs' textual outputs are meaningfully grounded in multimodal content. My response is that they would thus be grounded in a manner unlike that of any past participant of the space of reasons. Perhaps LLMs do possess agency through supervised learning. My response is that this process of learning is of a kind that no participant in the space of reasons has ever undergone. At a certain point these differences add up, and the agent "giving reasons" is not participating in the same game because they are not interacting with the same shared objects and the same shared practices as the other players.

### 4.5 – Conclusion: Pragmatism and the Normative Status of LLMs

There are several objections that could be posed against the above account. The most pressing is that I have merely shifted the goal posts such that my reconceptualization of the space of reasons undermines its explanatory power. I think this worry is a legitimate one. But I believe the benefits of my recontextualization outweigh the losses, especially if one is suspicious of the idea that the space of reasons is more than a useful concept, that it captures transcendental conditions for thought. I thus want to restate the practical benefits of my position in three ways.

First, as I have argued in Part I, I hold that my reconceptualization of the space of reasons trades Brandom's narrow formalism for a broader conception of when and where the game of giving and asking for reasons takes place. My position would not exclude any persons who are already said to occupy the space of reasons, nor does it necessarily exclude future artificial participants. Instead, the account preserves some implicit component of reason-giving that has been excluded from prior accounts, namely that reasons need not be and often are not linguistic performances. I therefore maintain that my position expands the explanatory power for the concept by providing a provision which limits certain kinds of artificial participants while also expanding the scope of possible justificatory performances that can be tracked or analyzed.

Second, my position captures the notion that LLMs remain a proprietary technology of firms who should bear responsibility for the errors produced by these systems. An identifying feature of artifactual normativity is that norms emerge within social and embodied practices

through intersubjective sanctioning of proper and improper uses of artifacts. Given this idea, Sambrotta (2025) is correct that the training undergone by LLMs does not allow for continuous updating of practices which results from sanctioning, undermining their normative autonomy in the space of reasons. However, I think this issue extends even deeper. Even if LLMs could engage in scorekeeping with interlocutors, this is already a superficial level of normative engagement. To be capable of effectively navigating the moral space, LLMs would need to be embedded in the moral practices themselves. Otherwise, they are merely repeating or stating the norms of practice without participating in them. Therefore, LLMs' lack of participation in a field of practices bars them both from standard modes of reasoning and from possessing a position within normative practice.

Third, my position does not bar future embodied AI from participation and therefore, while it possesses some anthropocentric qualities (as almost all accounts of reasoning do), it does not restrict certain forms of artificial intelligence from being in the space of reasons. I maintain that it is possible for future artificial agents to possess normative or moral agency which renders them, rather than a third party, the object of normative sanctions and scrutiny. One can think of various science fiction examples—Data in *Star Trek*, R2D2 in *Star Wars,* or Bender in *Futurama*—where non-sentient artificial agents possess a robust, embodied autonomy where they occupy a position in the world which is not derived from their manufacturer's intentions.[87] But as far as current progress in embodied AI and robotics is concerned, I think we are far from these fictional scenarios.

What these three qualifications suggest is that, going forward, the pragmatist, as a good conceptual engineer, may have to break apart and reconfigure the general concepts that they have utilized to discuss issues of intelligence and reason. Brandom's notion of discursive intentionality binds together ascriptions of belief, rationality, meaning, and moral agency. But the problems posed by artificial intelligence may require dismantling this neat bundle, especially as it proves increasingly difficult to determine which externalist and intentionalist stances actually hold. As we come to describe the behavior of these systems using the intentional stance and ascribe emergent

[87] HAL 9000 is a more difficult case, since he lacks a standard body but is still situated in the world through several audio-visual terminals. See Dennett (1997) for a discussion on how this case complicates a strict embodiment criterion. However, since this example, like the others, is merely fictional, it does not pose a concrete objection to positions on embodiment.

inferential capabilities to them, perhaps we might have to give up some of these concepts in order to preserve the others.

I argue that our conceptions of rationality and moral agency are those which should be preserved. Granting moral agency to LLMs detracts from the social and political background in which they are designed and created, a background that they cannot transcend by traversing into an uncontrolled environment. And granting them access to the space of reasons obscures the plurality of possible modes of justification that we enact in everyday life. In giving up our concepts of moral and rational agency, we not only bring aliens into the space of reasons; we readily confine ourselves to an alien rationality. We become alien reasoners of our own making.

To avoid alien modes of reasoning, my view entails a reluctant agreement with those positions within the AI sapience debate that see LLMs as a new class of instrument that produces truth-evaluable content (see Freiman and Miller 2020; Arora 2024). On one hand, these views capture the novelly of LLMs while refusing to grant them reasoning capabilities or moral agency. On the other hand, my reluctance comes from the fact that these proposals all assume a model of reason-giving that is itself alien. They miss that the gulf between human epistemic agents and LLMs lies in our ability to effectively channel and direct the material of our environment for rational ends. This missing ingredient is more than incidental. As I will show in the next chapter, so long as the justificatory performances of the space of reasons are understood as only textual, propositional, discursive, sentential, or linguistic, we will continue to underappreciate the robustness of human practice and overestimate the intelligence of the machines.

# CHAPTER 5

## DEEPFAKE DETECTION AND ARTIFACTUAL DELIBERATION

> By conferring a means of self-delineation of objects, of "statement without syntax," photography gave the impetus to a delineation of the inner world. Statement without syntax or verbalization was really statement by gesture, by mime, and by *gestalt*.
>
> — Marshall McLuhan, *Understanding Media* (1964)

### 5.1 – Introduction

It has been claimed that deepfakes, or AI-generated audio-visual media, will cause devastating effects to political discourse, resulting from the unprecedented realism afforded by the latest advancements in deep learning (Rini 2020; Fallis 2020). Against this narrative, recent epistemological debates have shifted toward an anti-skeptical position on deepfakes which deflates broader geo-political concerns and instead highlights the non-epistemic harms posed by nonconsensual pornography and digital harassment (Habgood-Coote 2023; Harris 2021). These positions recognize that deepfakes generate harms, but they contest the novelty of synthetic media compared to earlier forms of disinformation and argue that overstating the discontinuity between deepfakes and other kinds of audio-visual manipulation may be theoretically counterproductive.

Drawing from the resources of artifactual pragmatism, in this chapter, I defend a form of the anti-skeptical position by highlighting the active role of artifacts within the practice of reason-giving. First, I will consider a counter to the anti-skeptical position. Popular presentations of deepfake technology, buttressed by empirical research into deepfakes, heavily emphasize the unprecedented realism of synthetic media and insist that computational methods of detection outperform human participants in deepfake detection studies. This research generates problems for the anti-skeptical account since it concludes that deepfakes might become undetectable even by experts with enhanced detection technologies. According to this narrative of the "epistemic arms race," it is only a matter of time before synthetic media undermine all possible avenues for verification, effectively negating the epistemic status that recordings once had.

Against this narrative, I advance my own anti-skeptical argument. I show that studies assessing the human detection rate for deepfakes overstate the case for epistemic pessimism about

recordings. These studies are designed in such a way that their subjects cannot rely on social deliberation with others. By framing human detection in such a narrow way and comparing these results to algorithmic detection methods, research in communication studies, human-computer interaction, and computer science promotes a narrative which dismisses social and institutional responses to AI disinformation.

Taking such limitations into account, I propose that we should consider responses to deepfakes which do not depend solely upon novel technologies. As an alternative to detection, I consider the possibility of *artifactual deliberation,* the notion that we engage in social deliberation *with* technological artifacts. Artifactual deliberation requires shifting our understanding of how deepfakes are verified. First, we should shift away from framing deepfake verification in terms of *detection* and towards verification as *deliberation,* where multiple social actors are able to express their views with and against each other. Second, we should consider how deliberation can be expressed through *multimodal argument,* where various kinds of media artifacts are actively deployed as evidence within social-epistemic deliberation. Together, these two positions are combined into a framework which skirts the line between the too-oft disparate fields of analytic social epistemology and media studies, charting out the potential for future cross-disciplinary research into synthetic media.

### 5.2 – The Epistemological and Ethical Harms of Deepfakes

Within the philosophical literature, the harms of deepfakes are primarily presented in relation to epistemological considerations of recordings and their status as evidence, with initial perspectives proposing a form of skepticism or pessimism against the reliability of recordings in the face of increasing deepfake proliferation. These perspectives can be divided into two kinds of argument. The first holds that recording technologies will be undermined in a manner that will have widespread effects to other areas of epistemology due to their *regulatory effects* on spoken and written testimony (Rini 2020). The second concerns the *probability* of fake recordings and how the increasing rate of fakes reduces the epistemic value of genuine recordings (Fallis 2020; Harris 2022; Matthews 2023).

Regina Rini (2020) gives the most important variant of the first line of skepticism. For Rini, video and audio recordings function as an *epistemic backstop* which regulates other testimonial norms (2). This is because they are used to provide acute correction for lies and false

testimony and also passive regulation of norms due to the consequences of potential correction (2). Despite playing a major role in "supporting the rationality of our testimonial practices" (2), Rini points out that the epistemic role of recordings has been almost universally neglected within analytic philosophy on the grounds that it is tied to historically contingent social practices rather than universal features of human cognition (4). As a result, Rini worries that we are unprepared for the rupture caused by AI on our epistemic practices. This rupture will come in the form of what Rini (2020) calls *backstop crises*. These crises are events which undermine the regulatory functions of recording technologies through the introduction of hyperreal fakes (7). Rini predicts that several backstop crises in a row will modify our attitudes and resultingly undermine *all* recordings (8). Not only will testimonial norms come undone, but because audio and video recordings are also epistemic backstops for photographs, the role of other potential epistemic backstops will also be nullified in a cascading effect (12).

Don Fallis (2020) advances similar skeptical worries, but his views are more tempered and gradual. Fallis posits that the presence of deepfakes causes recordings to carry less information (5). This is because according to mathematical models of information, a video carries information of its subject when its existence demonstrates a higher probability that the original event occurred than not (9). Fallis argues that given this definition of information, deepfakes reduce the amount of information that videos carry since they increase the probability of false positives (10).

Keith Raymond Harris (2022) and Taylor Matthews (2023) come to a similar conclusion in applying common epistemological theories to deepfakes. Harris (2022), in particular, argues for *a propensity view* regarding the epistemic effects of fakes (including deepfakes). This view holds that "it is the propensity of convincing and relevant fakes to exist in an environment that threatens the acquisition of warrant from legitimate information in that environment" (16). In other words, the presence of relevant alternatives undermines even the successful acquisition of knowledge (9-10). Matthews (2023) explains that this type of skepticism results from classic "fake barn" problems in epistemology. In Alvin Goldman's (1976) original "fake barn" problem, a subject correctly perceives a real barn despite unknowingly traveling through a county whose local hobby is to make fake barn façades. According to Goldman's proposed *reliabilist* position, the causal link between the man and the real barn does not constitute knowledge because the subject's perception was not really a *reliable* causal process as determined by the rate of success for correctly seeing a real barn. Drawing from this example, Matthews (2023) argues that deepfakes act as a digital

analog to the fake barns case, threatening to render unreliable our engagements with genuine sources of knowledge.

Though these effects are not as radical as those theorized by Rini, such a shift in reliability will have generally pessimistic political effects. While Rini fears that backstop crises might undermine *all* recordings and their potential to serve as checks for testimony, Matthews (2022) envisions deepfakes as having unfavorable modifications to our *character*. In particular, he worries that deepfakes will promote broader intellectual cynicism towards recordings which "will dispose its possessors to make inferential leaps that all unregulated, digital, audio-visual information online is potentially deepfaked" (76). A similar problem—the liar's dividend—occurs when individuals who are put in compromising positions by recordings appeal to deepfakes as a way to avoid responsibility (Chesney and Citron 2019).

While the earliest scholarship on deepfakes within philosophy predict a widespread skepticism would emerge in relation to the epistemic functions of recording technologies, a second wave of scholarship argues that the initial concerns were overstated. These include appeals to the robustness of our social-epistemic practices (Habgood-Coote 2023) and the context-sensitive features of deepfake deployment (Harris 2021; Atencia-Linaries & Artiga 2023) in order to undermine the abstraction of the appeals to skepticism. However, the new anti-skeptics do not fully downplay that some negative ethical and epistemic effects emerge from deepfakes, rather they may be said to endorse what David Chalmers (2022) calls a "limited anti-skeptical conclusion" (241), where the negative effects of deepfakes are recognized but the *novelty* of these effects is reassessed to avoid promoting the over-skepticism which enables intellectual cynicism and the liar's dividend.

Joshua Habgood-Coote (2023) has the most robust argument for limited anti-skepticism. Habgood-Coote (2023) targets a particular narrative present within the scholarly and popular discussions of AI which he calls the "epistemic apocalypse" (2).[88] This narrative is motivated by three subclaims:

1. Deepfakes will have terrible effects on our socio-epistemic practices.
2. Deepfakes are historically unprecedented.
3. The solutions to deepfakes are technological

[88] In many respects, this position tracks a broader critical narrative in the ethics of AI, one which resituates current social and political harms over presumed catastrophic and existential risks (see Powers and Ganascia 2020).

Against the claim that deepfakes will undermine our socio-epistemic practices, Habgood-Coote points to the social situatedness of the deployment of deepfake technology and the constructed norms of "documentary" photography within journalism which has been presumed to result purely from the mechanical process itself. This is not to say that there is no difference between recordings and handmade pictures, but rather, following the work of Sanford Goldberg (2012), Habgood-Coote argues that recordings, like instruments, are artifacts "designed to track features of the world, whose reliability depends on the way they are designed, operated, and maintained" (7). In other words, just as the reliability of a mercury thermometer is dependent upon applying the correct measurement lines within the production process and ensuring in quality control that there are no defects in the produced output, so do recording technologies require institutional norms to manage their reliable properties.

Habgood-Coote (2023) uses this insight to then undermine "techno-solutionist" narratives which frame the epistemic problems of deepfakes in a manner that avoids addressing the social practices in which the epistemic robustness of recordings is maintained. Habgood-Coote's (2023) response is thus to out that the "techno-solutionism" framing has obscured the history of deepfake deployments, where the cited cases of deepfakes used as political disinformation were easily verifiable by looking at them and could be further analyzed by background verification occurring within our social practices (15).

Following this line of anti-skepticism, others have pointed out how the context-dependent sender-receiver relationship substantially affects the perceived unreliability or informational loss of recordings even if we are to *abstractly* assume that there is a higher propensity of deepfakes overall. Keith Raymond Harris (2021) outlines this position, pointing out that certain channels of information will be reliable regardless of any new potential for manipulation. For instance, deepfakes allow for a CNN broadcast featuring a well-known journalist to be faked and distributed over social media, but institutional norms would make it highly unlikely that one would find a deepfake broadcast distributed through the news network itself (2021, 133284). This undercuts skeptical arguments by showing that the overall propensity of deepfakes cannot be used to undermine the epistemic reliability of belief-forming processes in a variety of subcases.[89]

[89] Harris' (2021) position does lead to skeptical conclusions concerning the reliability of recordings outside of institutional sources, such as recordings posted on social media. Additionally, the question of reliability may still arise in contexts where institutional checks on news media are less robust or where the state has a monopoly on news media (see Schoenherr 2023)

Similarly, Paloma Atencia-Linares and Marc Artiga (2023) motivate the view that though the total evidential weight of recording technologies has been depleted, the actual effect of deepfakes on our epistemic practices is negligible. Turning to studies of animal signaling to explain how basic social and biological mechanisms lead to the upkeep of reliable signals, they posit four safeguarding mechanisms for signaling behavior in animals which can be reapplied to recording technologies. These mechanisms are:

> (i) existence of common interest (ii) developing signals that can only be faked with great difficulty as a result of limitations on the structure or design of the organism (indices); (iii) making the signal somehow costly to produce (handicaps) and (iv) punishing cheaters. (Atencia-Linares & Artiga 2023, 4)

Since we still have a vested interest in maintaining the reliability of recordings and sanction individuals for producing deceptive recordings, human forms of these four safeguarding mechanisms can be used to defend an anti-skeptical position with respect to the epistemic effects of deepfakes; there is reason to believe that our social practices can maintain the epistemic reliability of photographs despite technological threats to this reliability.

Though the limited anti-skeptical position rejects the narrative of the epistemic apocalypse, most scholars are careful to not dismiss deepfakes entirely. Though he criticizes the crisis framing in which deepfakes have been popularized, Habgood-Coote (2023) sees the negative effects of deepfake pornography on marginalized populations as a continuing threat. He adds that such depictions are harmful not because they produce epistemic harms but because they view women as "fungible subjects, whose faces and bodies can be swapped around like a children's toy" (19). This position echoes views from outside of philosophy which have primarily situated deepfakes in relation to problems of digital harassment and nonconsensual pornography. In *The Verge*, Russell Brandom (2019) provides an early public defense of the limited anti-skeptical position by appealing to actual use cases, concluding that "deepfakes are being used for misogynist harassment, not geopolitical intrigue." Scholars in media studies have also expressed disapproval with the general shift in coverage from concrete cases of exploitation to hypothetical political and epistemological crises (Burkell and Gosse 2019; Gosse and Burkell 2020; Maddocks 2019). Unlike political disinformation, deepfake pornography generates affective harms that cannot be mitigated through verification or correction (Maddocks 2019, 416). At the same time, the sexual and political deployment of deepfakes goes hand-in-hand, with nonconsensual pornography being used to threaten, silence, or harass political opponents, usually women and marginalized persons

(Maddocks 2019; Franks and Waldman 2019). This includes not only the targeting of individuals but the increased proliferation of racist, misogynist, or xenophobic stereotypes through synthetic media, some of which are made for purposes of promoting political extremism or reinforcing social bonds, rather than for deception.

### 5.3 – Deepfake Detection and the Epistemic Arms Race

Though I am sympathetic to the position that the problems of deepfakes are best mitigated through social and political interventions rather than through technological means, one central counter to the limited anti-skeptical approach is the sheer realism afforded by generative AI. The criticisms posed against the novelty of deepfakes masks the underlying empirical reality that deepfakes are becoming so realistic that there are no trace artifacts of manipulation within the recordings that could be used to make any sort of veridical assessment. To cite one early example, in 2019 *Mother Jones* reported on a New Years address by Gabon's President Ali Bongo, whose unusual appearance, combined with his reclusive behavior, sparked speculations of AI-manipulation (Breland 2019). The video was notable, not only because it was blamed for inciting a failed military coup against the longstanding regime, but also since various experts, including digital forensics pioneer Hany Farid and the fake video detection firm DeepTrace Labs, were unable to make determinate assessments on the video's authenticity (Breland 2019; Rini 2020).[90]

The worry that deepfake generation can outrun or outmaneuver current methods of detection plays into a major media narrative which runs parallel to and complements the epistemic apocalypse, one which I call the *epistemic arms race*. According to the epistemic arms race, synthetic media evade detection from all past verification methods in a manner that necessitates our reliance on technological detection methods.[91] This framing gamifies AI detection, treating the algorithmic operations of generation and detection as set in a competition or sport in which one side "wins out."

[90] The most plausible scenario is that the video was not doctored. Ali Bongo's odd mannerisms, different speech patterns, and lack of regular eye movement could be attributed to the long-term effects of a prior stroke. Bongo's subsequent public appearance the next year undermined rumors that he was incapacitated or dead. (Breland 2019; Rini 2020).

[91] I describe the phenomena as 'epistemic' to distinguish between the geo-political and military competition involving AI infrastructure and the competition between generation and detection algorithms which relies upon a similar war metaphor. My definition also tracks an arms race for text-based generative AI, but I will focus solely on audio-visual media for the remainder of the article.

The epistemic arms race was a common fixture of popular presentations of deepfakes in the first year of mainstream press coverage and it remains a common metaphor to the present day. In addition to framing generation and detection as functions of competing algorithms, the epistemic arms race also situates humans as defenseless against the threat of deepfakes, their "naked eyes" requiring supplementation by the latest advancements in machine vision. Additionally, early accounts especially noted how the epistemic arms race was born out of a more literal AI arms race, one which invokes a concern for the security of nation states. In particular, the US government's involvement in deepfake research in 2018 sparked such framings. As William Knight (2018) reported for the *MIT Technological Review*: "The Department of Defense is funding a project that will try to determine whether the increasingly real-looking fake video and audio generated by artificial intelligence might soon be impossible to distinguish from the real thing—even for another AI system" (para. 2). This framing of detection in terms of competing AI systems, a generator and detector, documents a real concern regarding the "detection dilemma," where the greater accessibility of AI detector is inversely related to its effectiveness (Leibowicz et al., 2021). This problem stems from the underlying mechanisms of artificial neural networks within generative models which can be trained to override new detection efforts as they arise.[92] The effect is "a constant seesawing of defensive and offensive advantage" (Engler 2019).

To make matters worse, gamified deepfake detection tasks show users how ineffective they are at distinguishing real images from synthetic ones. A popular *BuzzFeed* video featuring comedian Jordan Peele as a digital impersonation of Barack Obama was intended to trick unsuspecting viewers, ending with a warning about the dangers of manipulated content (Peele 2018). Likewise, a video by *The Economist* on the harms of deepfakes poses a detection game to its audience, concluding with the statement "one of the interviews in this film was a deepfake. But which one?" (The Economist, 2019). An article in *The Guardian*, takes a more proactive approach, highlighting the visual clues one can use to identify a deepfake, such as poor rendering, lack of image detail, and lighting discrepancies (Sample 2020), and the *BBC,* in a more recent profile of digital forensic firms, refers users back to a "Deepfake Quiz" where they may test their detection abilities (Wakefield 2024).

[92] The first deepfakes were powered by Generative Adversarial Networks (GANs) consisting, in part, of a generator and a discriminator which competed to produce a more realistic output. For the original paper on this neural architecture, see Goodfellow, et al. (2014).

These deepfake detection tasks have extended into academic presentations of deepfake detection. In a mixture of public engagement and scholarship, whichfaceisreal.com, a website hosted by University of Washington professors Jevin West and Carl Bergstrom, shows users two still images of faces and prompt them to "click on the person who is real." The stated goal of this detection game is to increase digital literacy with respect to identifying the visual artifacts left by generative AI. However, despite their positive intentions, West and Bergstrom (2019) remain skeptical of their project's goals:

> If you know what to look for, you can spot these fakes at a single glance — at least for the time being. The hardware and software used to generate them will continue to improve, and it may be only a few years until humans fall behind in the arms race between forgery and detection. (para. 5)

While this kind of detection task informs audiences of the increasing realism of synthetic content, they do not demonstrate that deepfakes are actually increasingly more difficult to detect by humans in a manner that renders them unprecedented.

To fill in this gap, comparative studies assessing human detection ability have emerged within communication studies and human-machine interaction. One study conducted by Vaccari and Chadwick (2020) tested audiences' detection abilities through three sets of participants – two sets individually watched the false portions of *Buzzfeed's* Obama deepfake video while another control group engaged with the full video. The results showed that only 50.8% of test subjects were not deceived in some way (Vaccari and Chadwick 2020, 6). Other studies have used a comparative paradigm to test human subjects versus computational detection methods.[93] One study assessing the effects of image quality on detection, found correct human detection rates to be "68.69% in average on raw videos, 66.57% on high quality, and 58.73% on low quality videos" which was lower than the rates of computational detection methods (Rössler et al. 2019, 5). Test subjects in a September 2020 study were deceived at a significantly higher rate, with 75.5% of participants struggling to determine which videos were false, although the comparative advantage of humans versus machines was less clear (Korshunov and Marcel, 2020). Finally, a newer detection study shows that humans score only slightly better than random in their detection abilities

[93] A variety of methods have been proposed, with initial proposals including tracking eye-blinking rates (Li et al. 2018), heart rhythm patterns (Qi et al. 2018), and inconsistent head poses (Yang et al. 2018). For some recent surveys on computational detection methods, see Mirsky and Lee (2020), Taeb and Chi (2022), and Sharma et al. (2024).

(Bray et al. 2023) with interventions to educate study participants having no effects on outcome (Bray et al. 2023; Somoray and Miller 2023).[94]

The implications of this line of research pose a major threat to philosophical arguments against epistemic skepticism about deepfakes. The framing of the epistemic arms race hints at a future where no human being or instrument could ever verify the veracity of a recording, thereby complicating anti-skeptic appeals to social practice and context-sensitivity, for even journalists, digital forensic experts, and the state-of-the-art detection algorithms may fall victim to an undetectable fake. If we take the epistemic arms race narrative seriously, perhaps there already are a sea of fakes populating our social networking sites which will never be discovered as such. Perhaps our epistemic environment is already polluted by undetectable deepfakes. Even worse, even if detectors could be rendered more robust for long-term applications, the damage of deepfakes is already done. For in the optimistic scenario, our reliance upon detection technologies would outsource important human capacities, reducing our overall epistemic autonomy (Harris 2024). The future then is either a pure hyperreality where truth and fiction are indistinguishable or one of perpetual technological dependence for basic informational needs.

### 5.4 – Artifactual Deliberation

Though the notion of the epistemic arms race captures the basic intuition that synthetic media could, in principle, avoid any method of detection, there is also a sense in which this framing fails to address the concrete practices of verification. On one hand, the detection paradigm correctly alerts us to the fact that deepfakes may readily overwhelm the reliable channels of information which are often appealed to as a defense against synthetic media. It suggests that the appeal to powerful government institutions, well-meaning journalists, and qualified forgery experts does not adequately address the problem of undetectable fakes, since these channels may themselves be unable to assess a recording's veracity. On the other hand, the anti-skeptic correctly counters that deepfake research often naively presumes that verification means "detection" and that detection is always *immediate*. These presumptions afford greater experimental control to the researcher. But theoretically, they risk artificially compartmentalizing human test subjects in a manner that undermines human epistemic autonomy and overstates epistemic harms.

[94] Additionally, Birrer and Just (2024) provide a comprehensive survey of empirical research into human detection rates.

Vaccari and Chadwick (2020) effectively summarize the limitations of the detection paradigm when they note that their own experiment "did not account for interpersonal networks, algorithmic filtering, and competition with other messages—all factors that are likely to play a role in promoting or debunking deepfakes outside the contained environment of an experiment" (Vaccari and Chadwick 2020, 8). This admission is notable since it challenges the ability to extrapolate detection studies to most real-world cases, but it does not undermine or downplay the importance of empirical research as a tool for assessing the differences between AI disinformation and earlier forms of audio-visual manipulation. Disinformation research can track the applicability of the "realism heuristic" (Sundar 2008) and other proposed medium-specific effects on disinformation uptake, such as differences that audio-visual modality plays in belief modification relative to other modalities (Hameleers et al. 2022; Lee and Shin 2022). Where this line of research falters is in making socio-political inferences on the basis of psychological effects.

I propose that an anti-skeptical appeal to social practice can help generate new paradigms which more effectively map to actual practices of verification. My position is that deepfake verification occurs through the process of *artifactual deliberation*, where participants in the space of reasons engage in social deliberation through a variety of linguistic and non-linguistic artifacts. Artifactual deliberation responds to the methodological limitations of the detection paradigm by integrating two theoretical accounts. First, it sees detection as tracking individual moments within a broader process of *social deliberation* where verification is a dynamic, inter-subjective practice of reason-giving. Second, it incorporates *multimodal argumentation theory* in order to account for the variety of media that may be utilized in social verification. Together these two accounts are combined to track how participants within the space of reasons rely upon the epistemic affordances of situational artifacts to buttress linguistic conversational norms and how these conversational norms, in turn, render artifacts available for use as reasons. Hence, my broader thesis is that the capacity to deliberate *about* technologies becomes enmeshed with a capacity to deliberate *with* them.

### *5.4.1 – From Detection to Deliberation*

Some empirical studies into deepfakes have recognized the overreliance on methodological individualism in detection tasks and have tried to amend their treatment of detection. In an appeal to the "wisdom of crowds," Groh et al. (2021) found that the aggregate mean score of individual

assessors was equal to or even better than computer vision models, demonstrating that when detection is considered as a *group task*, the arms race narrative is substantially weakened. Yet even here, only the group was assessed according to their correctness of algorithmic scoring without any additional connection with other participants. While Groh and colleagues mitigated some of the methodological problems of the detection paradigm, they still failed to capture real-world practices of verification and therefore, cannot be taken as providing a model for assessing deepfake harms. One way to understand the remaining methodological issues is to look to deliberative accounts within social epistemology.

In particular, Elizabeth Anderson (2006) provides a helpful model which tries to capture the epistemic elements of an open democratic society. Detailing her pragmatist and experimentalist social epistemology, Anderson appeals to three constitutive features of an epistemic democracy, i) *diversity* – it includes several perspectives or viewpoints, ii) *discussion* – it involves interaction between these viewpoints, and iii) *dynamism* – it features self-correction or modification over time (13). When we consider deepfake verification, it is important to consider these three features, that human viewers bring with them different conceptual backgrounds and political-strategic goals which bear upon image detection tasks (diversity), that they draw from their different conceptual backgrounds to converse with others (discussion), and that, while some individuals might be initially fooled by a deepfake, there is always the possibility for correction overtime (dynamism).

While the "wisdom of crowds" study correctly leveraged the effects of numbers (Groh et al. 2021), it did not capture any of the three features of democratic deliberation. First of all, since each individual respondent is isolated from every other respondent, the study is more properly understood as a variation of the Condorcet Jury Theorem. This theorem states:

> If voters (a) face two options, (b) vote independently of one another, (c) vote their judgment of what the right solution to the problem should be (i.e., they do not vote strategically), and (d) have, on average, a greater than 50% probability of being right, then, as the number of voters approaches infinity, the probability that the majority vote will yield the right answer approaches 1. (Anderson 2006, 10–11)

The Condorcet Jury Theorem appeals to numerical decision-making, but it does not capture the *diversity* feature of democratic deliberation. Diversity requires rejecting conditions (b) and (c), since it predicates that individuals will form associations with other similar individuals and may vote strategically along these group lines. In the "wisdom of crowds" study, diversity was actually purposely selected against by the experimental design. As the authors note, "We limit the bulk of

our evaluation to uncontroversial videos of unknown actors to focus on the visual component of truth discernment" (Groh et al. 2021, 8). This essentially prevents actors from relying upon prior biases and expectations of known actors to figure into their assessments of veracity.

By excluding these kinds of associations, Groh et al. (2021) attempts to counteract discrepancies between human vision and machine vision in deepfake detection research where it was found that humans use conceptual resources about who and what is presented in a manner unavailable to computational detection technologies which instead analyze patterns of pixelation independent of the conceptual content presented within the images (Rössler et al. 2018; Goh et al. 2024).[95] These differences in visual processing have been further demonstrated in the documented failures of deepfake detectors against obvious fakes, such as of an image depicting a large Yeti-like creature (Thompson and Hsu 2023), where nearly all human viewers would be able to rely on conceptual information about the non-existence of Yetis to assess the image's veracity independent of image quality or realism.[96]

But at the same time, it is important to recognize that diverse conceptual backgrounds are ubiquitously relied upon in real-world social environments. Take for instance, recordings of fictional events which have been staged rather than generated (see Atencia-Linares 2012). In these instances, individuals must utilize background conceptual knowledge to compare the contents of the recording to other sources rather than rely upon the realism of the visual depiction. Ironically, when we consider the conceptual features of assessment, AI disinformation actually has more avenues for detection compared to staged recordings or technologically-simpler "cheapfakes" (Paris and Donovan 2019), which cannot be as readily verified through detection technologies.

This reflection lends itself to anti-skepticism. Consider the repeated example of hypothetical deepfakes that depict US soldiers committing atrocities abroad (Chesney and Citron 2019; Fallis 2020). One might ask: What is preventing someone from dressing up individuals in military uniforms and recording them supposedly performing these actions? Would this latter scenario be any more dangerous than that of a deepfake? Would it be more difficult to refute?

[95] See also Alvarado's (2022) discussion between AI expert systems versus human radiologists for a philosophical treatment of this problem.

[96] A method approximating the conceptual background of human viewers can also be run on multimodal large language models by representing objects in images as propositions and then checking these propositions against background data, as I had mentioned in Chapter 3. This alternative method of AI-verification is in development but fairly ineffective given the lack of up-to-date search features in MLLMs (see Chandrasekar and Jaźwińska 2025).

Would it draw the same level of academic and industry concern? In the spirit of the anti-skeptical position, we must consider whether hypothesized deepfake scenarios actually identify what is truly novel about AI impersonations, and the first place to start is a reevaluation of the presumed methodological individualism of the detection paradigm. When one understands that the relevant practices of verification involve the epistemic features of diversity, discussion, and dynamism, the novelty of deepfakes wears away.

### *5.4.2 – Multimodality*

As stated thus far, deliberative social epistemology still does not capture the concrete details of verification. One remaining problem lies in the question as to what counts as a move or performance within the space of reasons. In empirical domains, the prevalence of deepfakes has reinvigorated earlier research into *multimodality*, where the presentation of information through one or more forms of media has potential consequences for the uptake of such information (Sundar 2008; Hameleers et al. 2020, 282; Lee and Shin 2022). Even with this resurgence, visual disinformation is often neglected in research and theory with negative implications on our understanding of democratic processes (Weikmann and Lecheler 2023, 3697). In relation to a deliberative social epistemology, multimodality is important in that it challenges the presumed linguistic or verbal dimensions of communication by documenting alternate forms of rational persuasion which have been neglected within philosophy.

For example, when we look to the concrete instances of verification, we find that journalists do not rely solely on digital forensics tools, nor do they merely *talk about* visual media. They also use *circular verification strategies* and *background verification strategies*, respectively referring to comparative sourcing and to verification by tracing the history and origins of a particular source (Lecheler et al. 2019). In deliberative settings, these strategies could be enacted as "multimodal rebuttals" to visual disinformation (Hameleers et al. 2020). A face-swapped deepfake could thus be refuted by comparative analysis with a prior posted recording, and a full digital performance might be debunked merely by tracking its historical appearance to various disreputable sources. Digital literacy strategies involving reverse-image search may aid these forms of comparative and historical analysis (Qian et al. 2023; Weikmann and Lecheler 2024), providing quicker methods for tracking the diffusion of visual content.

These types of justificatory performances are difficult to theorize philosophically, perhaps especially so for the deliberativists to whom I appeal, given the association of deliberation with communication. These performances require the presupposition of a distinct kind of non-linguistic, non-perceptual evidence which is brought into our intentional and persuasive activities. As I argued in Chapter 3, when one shares a photograph, the justificatory performance is neither strictly perceptual nor testimonial. Instead, I argued that it is properly understood as an artifactual reason. Much of the philosophical literature on deepfakes has been indirectly concerned with how to parse out knowledge-by-recording within the traditional debates on perception versus testimony and inferential versus non-inferential knowledge (see Cavedon-Taylor 2013; Rini 2020, 9; Cavedon-Taylor 2022; Habgood-Coote 2023, 103). Artifactual reasons are a means to bypass this debate by identifying the justification as occurring in the use of a particular epistemic artifact. However, only recently has the philosophical conversation begun to frame the topic of audio-visual disinformation away from the perceptual-testimonial divide and toward the medium in which an assertion or justification is made.

Two recent papers have enacted this shift by appealing to multimodality. First, from a Gricean standpoint, Megan Hyska has asserted that environmental objects can be embedded within what she calls "multimodal speech acts" (2005, 141), arguing that such a theory can be applied to demonstrate how recordings lose status as acts of *showing* and become instances of *tellings*, where the sender's communicative intentions play a primary role in the acceptance of an assertion rather than the affordances of the world being deployed within the assertoric act. Hyska's intervention highlights how deepfakes threaten to undermine the weight that recordings have as evidence for marginalized persons and communities, where perceived credibility deficits of epistemic agents could once be supplemented with recordings.[97] A similar appeal to multimodality comes from the standpoint of informal logic, Liao and Niu suggest that multimodal and visual argumentation theory may be fruitfully applied to map instances of image forgery and deepfake proliferation, especially when supplemented by research in media and communication studies (2025, 28).

These positions echo longstanding efforts within argumentation theory to extend the constituent parts of rational persuasion to include alternative, non-verbal modalities. Many such

[97] Maddocks (2019) argues that the use of deepfakes to silence of women and the subsequent recentering of political harms away from pornographic abuse constitute forms of "epistemic injustice" (see Fricker 2007). For an overview of epistemic injustice in AI, see Symons and Alvardo (2022)

efforts can be traced back to a series of defenses for the possibility of *visual argument* in the mid 1990s, in particular, Leo Groarke's (1996) position that visual media, such as pictures and photographs, possess a collection of assertions or statements which "like a collection of verbal statements, function as reasons for a conclusion" (111). The details of Groarke's account have sparked a series of exchanges which can be mapped according to three contested aspects of multimodal argument: 1) the *normative criteria* for non-verbal modes of argument, 2) the definition of a *mode* or *modality*, and 3) the *propositional contents* of non-verbal modes.[98]

To briefly recount each aspect of the theory for the purposes of elucidating my own view, the first aspect of multimodal argumentation revolves around revision to the underlying *norms* of argumentation. Drawing upon Brandom's work, David Godden, a moderate defender of multimodalism, sees this as an issue regarding whether there is one distinct, logical *space of reasons* in which many modes of argument participate or whether each mode possesses its own separate space of reasons with its own assessment criterion (2022, 525). The first model is compatible with what Godden calls a "trans-modal evaluative non-equivalence," which argues that premises expressed in one mode may have a different rational or probative assessment compared to other modal expressions (540). The more radical view, what Godden terms "normative revisionism," argues that the very standards for evaluating an argument change between modalities, with non-verbal arguments requiring an entirely different logic from verbal ones (541).

The problem of deepfake verification heavily aligns with the position of trans-modal evaluative non-equivalence but not with full normative revisionism. This can be seen in both Rini's (2020) position on the epistemic backstop and anti-skeptical responses documenting the socio-technical practices surrounding recordings (Habgood-Coote 2023; Atencia-Linares and Artiga 2022). Both sides of the debate presume that recordings have a different reliability or informational capacity compared to text-based sources, but neither side holds that deepfake assessment requires new evaluative standards for argument. This convergence in views might be expressible in terms of premise acceptability (Liao and Niu 2025, 28); both sides of the deepfake debate share the position that premise acceptability may change depending upon the vehicle in which a premise is asserted, but these changes to premise acceptability do not change the standards for ascertaining valid premise-conclusion forms from the overall argument.

[98] These three aspects are my own, but they follow the similar treatment provided by Godden (2022).

A second important aspect of multimodal argumentation involves the very status of a *mode*. Here, Godden (2022) breaks down the definition of mode into informational and materialist accounts. The *informational account* sees different modalities as expressing different degrees of information, with visual images expressing more information than words and video expressing more information than still images (Godden 2017). In contrast, the *materialist* account views different modes as different physical objects in the world, they are like different construction materials which interact to support and stabilize each other in a building (Groarke 2015, 140). In relation to deepfakes, both definitions of modality are prevalent (and they may even be complementary). The informational account tracks information theoretic perspectives regarding the more expressive content of recordings (Fallis 2020; Atencia-Linares and Artiga 2022), and the materialist account maps onto positions which appeal to the physical properties of video or to its affordances as an instrument or actant in our practices (Habgood-Coote 2023; Weikmann and Lecheler 2024).

The third central aspect of multimodal argumentation is an account of how non-verbal contents can be expressed as reasons. This aspect is the most contested one, upon which the entire edifice of multimodal argument relies. On one hand, opponents to multimodal argument endorse what I call an *irrationalist view of pictorial content*, which states that pictures are not propositional, and since only propositions can be reasons, pictures cannot be reasons. This argument is roughly the one given in analytic pragmatism by Davidson (1978, 263), Rorty (1989, 18), and Brandom (2009a, 220–221), as discussed in Chapter 3. But though these earlier positions may seem dismissive of non-linguistic features of argument, Alan Gross (2009), in a more charitable account, defends the autonomy of visual communication while at the same time rejecting the possibility for visual argument. For Gross, there are four essential differences between visual and verbal modes which prevent the visual from being propositional:

1. the verbal is processed sequentially; in contrast, images are processed not only sequentially, but also in parallel and simultaneously.
2. in the case of natural and artificial languages, internal connections exist among their fundamental components. In contrast, in the case of images, contiguity rules: they and their components are organized spatially into synchronous hierarchies or nested sets.
3. words never lose their individual identities; in contrast, the identity of images can easily be submerged in larger wholes, as eyes and ears in a familiar face.
4. unlike words, images can undergo meaningful spatial transformations and manipulations, such as superimposition, projection, rotation, magnification, and animation. (Gross 2009, 149).

According to this division, images are analyzable as Gestalt figures where part and whole blur together, while sentences have clearly demarcated subjects and predicates which reduce ambiguity (Gross 2009, 153). Hence, images are non-propositional by the very nature of their presentation.

Taking up a defense of Brandomian pragmatism, Frápoli (2023) equally dismisses multimodal argumentation theory on the grounds that visuals cannot, themselves, be propositional. Frápoli argues that most defenders of multimodality confuse the notion of *propositional content* and the *vehicle of representation* (176). According to rationalist pragmatism, propositions are logical entities which have no relation to the physical make-up of their vehicles, or as Frápoli puts it, "Truth, consequence, and incompatibility are properties of abstract entities or groups of them, and not of their linguistic representations" (186). This point may initially allow for visuals to be propositional in the sense that visual vehicles can be expressed as propositional contents, but Frápoli additionally argues that only sentences can make propositions explicit, and therefore, no visual contents can serve as premises in an argument without linguistic articulation (196).

Though Frápoli is skeptical of multimodality, her argument closely follows a moderate position which argues that visuals can, themselves, be propositional in the sense that visual vehicles are expressible as propositional contents. I call this the *propositionalist view of rational pictorial content*. Broadly, this line of argument holds that pictures express propositions, allowing them to serve as reasons, and it features prominently among defenders of multimodality (Blair 1996, 26; Birdsell and Groarke 2007, 106; Grzankowski 2015). However, the propositionalist is put into an uncomfortable position where they must defend how the part-whole structure of images can be made propositional. At best, this requires "dressing" an image to make its content explicit and standardized (Groarke 2015, 135).

In some sense, the disagreement between the irrationalist and the propositionalist might be semantic and rhetorical. One might argue that the disagreement lies in whether visual arguments can exist without sentences, but although there were some early adopters of this view, this is a position that most multimodal theorists have rejected (Blair 2014, 218; Groarke 2015, 134). The disagreement therefore lies in whether we should see visual propositions as items which are *made explicit* through language or *created* by language. Here, I think the propositionalist's use of "dressing" merely points out that there are a range of potential propositions to be asserted by a picture and that language helps us pick out the relevant ones. On the other hand, Frápoli (2023)

sees dressing as a conceptual processing of non-conceptual information (195–196), which suggests some radical change has taken place within the picture, that it has been placed into a broader grid or matrix of linguistic intelligibility, it has been "propositionalized."

Given these confusions, consider a third view that I call the *non-propositionalist view of rational pictorial content*. Like the irrationalist view, this position argues that pictures are themselves non-propositional. However, it also argues that non-propositional contents can be reasons (see Roque 2015; Godden 2017; 2022). In many respects, this view is revisionist about the definition of a reason; it is interested in generating a better concept of 'reason' rather than relying upon the standard philosophical assumptions. As Godden (2017) puts it, "To simply stipulate, or insist, that arguments consist of propositions and that pictures are nonpropositional is to beg the question against the [multimodal] enthusiast" (419). Roque (2015) adds that, given the changing status of our understanding of non-propositional concepts in the cognitive sciences, there may be good empirical reasons to reject the assumption that reasons must be propositions. As a proposal for a modified conception of reason, Godden (2017) appeals to *argumentative content*—"content that constitutes, or ineliminably contributes to, a reason or claim" (421). Following Brandom, this notion of content is defined *functionally* by how some content figures in the game of giving and asking for reasons, but since argumentative content replaces that of propositional content, any modality can participate in the space of reasons without being propositional so long as it contains material from which valid inferences may be derived.

### *5.4.3 – Multimodality and Artifactual Reasons*

In understanding these three aspects of the multimodal paradigm, my position on *artifactual reasons* may be recontextualized as a theory of multimodal argument which defends *normative non-revisionism* regarding the evaluative standards of argument, *materialism* regarding the status of modality, and the *non-propositionalist view of rational pictorial content*. In Part I, I argued from a contemporary pragmatist perspective that such things as photographs, signposts, slaps, and kisses can be brought into the space of reasons in a manner that alleviates the metaphysical concern that rational justification spins "frictionless" above the world. My position, artifactual pragmatism, preserves a singular, logical conception of the space of reasons, yet argues for a separate space of practice, which avoids conflating the properties and social norms of objects and artifacts with the norms of valid inference within the game of giving and asking for reasons.

At the same time, externalizing McDowell's position on perceptual content, I allowed *conceptualized* vehicles from the space of practice to be available for deployment as *reasons* without conflating concepts or reasons with propositions (non-propositionalism). Unlike McDowell, my preference for the terms 'pictorial content', 'conceptual content', or 'visual content' as opposed to 'propositional content' is not due to an understanding of Kantian intuition as a non-propositional grounding for judgment, but rather due to an anti-representationalist rejection of the content-vehicle distinction, a skepticism I share with both Dewey and Rorty.[99] While a Brandomian might insist that there are such things as "propositional contents" which can integrate any kind of vehicle (Frápolli 2023), I argue that the world is vehicles all the way down.[100] Sentences are strings of marks or noises which have been combined to form more complex artifacts. These artifacts have the capacity to make non-sentential conceptual contents explicit, but this capacity is merely an *affordance* of sentences as another kind of artifact. In other words, while a representationalist view claims sentences uniquely subtract the irrelevant properties to mirror hidden propositional structures in non-sentential things or that they "propositionalize" non-conceptual contents, the anti-representationalist maneuver is to claim that sentences *add* further properties to an already conceptually-articulated world in a manner no different from any other artifact; there is no recourse to facts as pre-human logical states of affairs nor to 'propositions' as abstract objects. The reason-cause distinction can be maintained according to the distinction between artifacts that express argumentative content and artifacts which do not.

Methodologically, these arguments suggest that the linguistic pragmatist's conceptions of normativity, conceptuality, and rationality are mistaken in a manner that fails to track concrete practice. One issue—also discussed in the prior chapter in relation to multimodal LLMs—is the disregard for audio-visual concepts which cannot be reduced to linguistic training. For example, the ability to infer from the statement 'JFK was assassinated' to 'A US president was assassinated' is different from the ability to infer from the visual contents of the Zapruder film to the belief 'JFK was assassinated'. Hence, the objects of what Brandom calls language-entry and language-exit

[99] Despite using 'content' in a different manner, the content-vehicle distinction looks like another form of the scheme-content distinction.

[100] I find that the materialist position on modality can endorse the non-propositionalist position of pictorial contents when it holds that all vehicles are held together within other vehicles. This view shares features with McLuhan's (1964) medium theory, in which more complex media can be broken down into simpler mediums. "Vehicles all the way down" is perhaps my more philosophically acceptable substitute for "the medium is the message."

transitions must themselves be items which figure into the available conceptual background of social-epistemic agents, items which can functionally serve as evidence between two agents who share a similar-enough conceptual background (both verbally and visually).

In outlining this feature of artifactual reasons, I need to demonstrate how this position uniquely captures practices of endorsement and rebuttal of visual media, practices which constitute a visual argument *without* linguistic dressing. First, as a caveat, I had mentioned in Chapter 3 that an account of artifactual reasons does not argue against the notion that participants in the space of reasons must have, in some sense, discursive capacities, and to my knowledge, none of the multimodal theorists outlined above take up such a position. The possibility of a purely visual argument is just the possibility that participants, who have discursive capacities or who *could* engage in verbal argument, can also engage in practices of reason-giving using visual media without relying upon immediate linguistic premises.

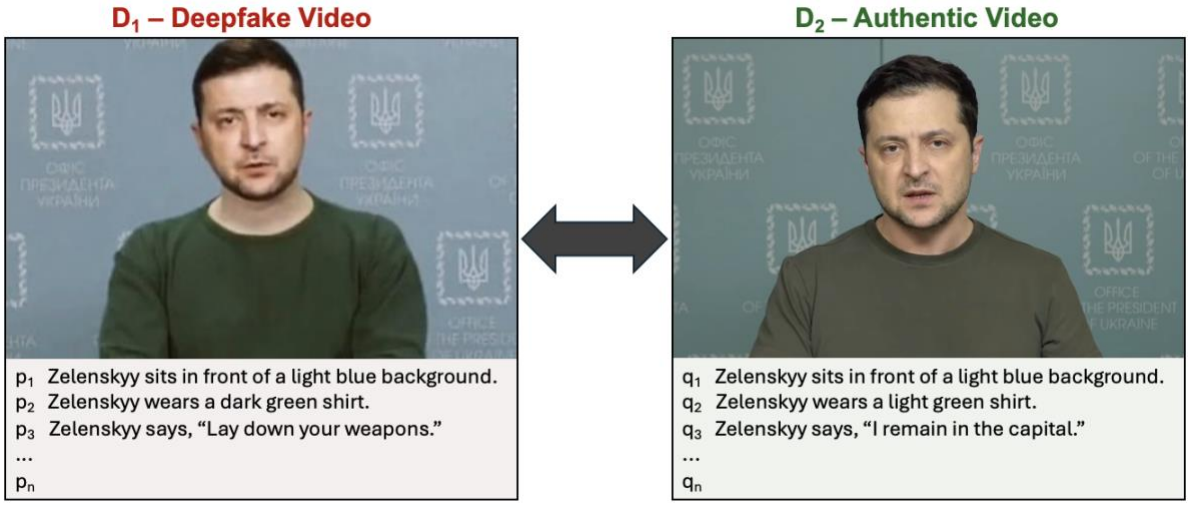


Figure 3: A comparison of a deepfake video with a corresponding authentic video.[101]

Consider the following scenario shown in Figure 3: Person A posts on a social media site a deepfake $D_1$ of Ukrainian President Volodymyr Zelenskyy. The video shows him telling his troops to lay down their weapons and surrender to Russian forces (Tanno 2022). This video contains a Gestalt part-whole arrangement of available visual contents which would lead a suitably

[101] Source images of the deepfake are from Tanno (2022) and the authentic video is sourced from Zelenskyy's official Facebook page (Volodymyr Zelensky, February 26, 2022). Images have been modified from their original size to emphasize similarity.

trained interlocutor to assert the set of sentences [$p_1$ , $p_2$ , $p_3$ , ... $p_n$]. Nonetheless, no such sentences are explicitly asserted. In response, imagine that Person B posts an unaltered video of Zelenskyy, $D_2$, with visual contents that would lead an interlocutor to assert [$q_1$ , $q_2$ , $q_3$ , ... $q_n$]. In this video, Zelenskyy instead says "I remain in the capital" (Reuters 2022). Upon closer inspection, an inspection sparked by $D_2$, most viewers see that $D_1$ contains visual artifacts which suggest that it has been manipulated, and they see that $D_2$ contains no such artifacts.

I think this type of exchange, what Hameleers et al. (2020) call a *multimodal rebuttal*, demonstrates a non-propositional and non-verbal practice of reason-giving. In this exchange, neither A nor B has "dressed" their respective videos with linguistic garb to specify some particular feature of the video as a premise in an argument. Nevertheless, I think it is safe to say that the exchange has taken place in the space of reasons, that B's video $D_2$ has negated some of the conceptual content of $D_1$ which was intended to rationally persuade an audience. Importantly, this negation was not expressed in terms of sentences. In posting $D_2$, B implicitly endorsed a set of alternative visual contents which negated some of the visual contents expressed in $D_1$. My suggestion is that this negated content is not necessarily *propositional content*, since it is difficult to see, in propositional terms, what B's posting of $D_2$ explicitly asserts. According to the propositional view of pictorial content, $D_2$ would assert something like "The propositional contents of $D_1$ are false" or it might contain multiple propositions which negate the propositions expressed by $D_1$. Yet one problem posed by multimodal theory is whether the part-whole structure of visual content can in fact be "propositionalized" without linguistic dressing.

I think the non-propositional view of rational pictorial content correctly theorizes that $D_2$ negates the visual contents of $D_1$ in a manner that has auxiliary effects on the sentences that may be asserted in the space of reasons or the beliefs generated by viewers, but that this move is not propositional in itself precisely because visual contents are conceptually articulated in a different way. Person B is still giving a reason, but the reason is instantiated in the artifact, in the Gestalt part-whole structure of the visual contents, not in the sequential subject-predicate form of propositions. When the part-whole structure of $D_1$ fails to match that of $D_2$, the negations are expressed visually in the frame-by-frame mosaic of misalignments. Through $D_2$, B *shows* the falsity of $D_1$ without *saying* anything.

### 5.5 – Conclusion: The Deep and Shallow Harms of Deepfakes

The above scenario is one example of a non-linguistic justificatory performance that can be mapped out according to an account of artifactual reasons. My thesis is that when these types of performances are combined with features of diversity, discussion, and dynamism, the result is a notion of social-epistemic practice which draws extensively upon the theoretical needs of empirical scholarship, and which shifts the overall self-image of social epistemology toward that of media studies.[102] To return to the initial problem of the epistemic arms race, I have argued that the arms race narrative has overstated detection at the risk of understating deliberation. When one sees the verification of deepfakes as a practice of artifactual deliberation, many of the skeptical concerns toward deepfakes are deflated. Still, artifactual deliberation cannot solve all of the potential problems posed by deepfakes. It can, however, help determine which harms are worth worrying about.

First, artifactual deliberation shows that the depicted subject's publicity and social capital will affect deliberation. In past popular depictions, deepfakes have shown major geo-political figures or celebrities as targets of digital impersonation, but these figures have a broad public standing where most epistemic actors have some reserve conceptual resources about them (both visual and verbal) which can be deployed in deliberation and where many of these actors possess some stake in defending or opposing the depiction. These high levels of diversity and discussion are not present for depicted subjects who are known only locally to a small group of potential deliberators (the vast majority of the global population).

As a second consideration, artifactual deliberation is most successful when the epistemic environment features diversity, discussion, and dynamism. In this manner, different platforms may have different affordances which decrease the potential of deliberation.[103] A politically polarized website like 4chan will have low diversity compared to Facebook, and Facebook will have a lower potential for discussion and dynamism compared to a more open platform like Twitter/X but more than a messaging app like WhatsApp. Similar effects are also produced by the broader socio-

[102] For the reverse position, that media studies should draw from philosophical resources in social epistemology, see Godler and Miller (2020)

[103] For an overview of platform-level affordances, see Bucher and Helmond (2018). For an account of the epistemic effects of platform affordances, see Marin (2022). For an outline of platform regulation for both text-based and audio-visual synthetic content, see Fisher et al. (2024).

political climate where deliberation occurs, with more authoritarian countries having less adaptable epistemic environments through which deliberation could occur (Schoenherr 2023).

A third consideration is that the extent and type of manipulation will affect the available kinds of justificatory performances that epistemic actors can enact in deliberation. Face-swaps are examples of "local" manipulation, while a fully virtual performance effects the media "globally" (Millière 2022). Local manipulations can be verified through performances where a corresponding media source can be used to negate the contents of the fake source. However, global manipulations involving deep learning would leverage the effects of a realism heuristic while also limiting the potential use of artifactual reasons as rebuttals. This is not to say that global manipulations are "undetectable," but rather that it is harder to prove that a depicted event did not occur when there are no correspondingly real recordings for comparison. For synthetic media, the global manipulation case can be handled by detection technology. However, if the detection technology fails or if the disinformation is a fully staged scenario—a real video of a faked event rather than a fake video—then it is harder to see how deliberation will produce epistemically correct outcomes.

Finally, artifactual deliberation can only defuse harms produced from disinformation as an epistemic phenomenon. It therefore does not track instances of wrongdoing which cannot be corrected through verification. One dimension it cannot address is the increasing interpersonal dishonesty that results even from innocuous use of generative AI in art, media, and advertising (Flattery and Miller 2024). In cases of virtual performances in cinema, AI customer agents, and AI-generated viral content, instances of wrongdoing occur regardless of whether deliberating actors determine the veracity of the synthetic content and regardless of whether the content is particularly troubling or offensive.[104] A second more-pressing dimension is the longstanding use of deepfake technology to create non-consensual pornography, as first reported by Samantha Cole (2017). Given that the vast majority of past deepfake content has been pornographic in nature (Ajder et al. 2019; Gosse and Burkell 2020), there are harms which will not be corrected by deliberation. Even worse, attempts to determine the veracity of pornographic deepfakes may amount trading and reproducing demoralizing and objectifying depictions of subjects as artifactual

[104] I thank Tobias Flattery for pointing out that my approach glosses over cases of wrongdoing where the moral or epistemic consequences are minimal. I think he is correct to locate these instances of wrongdoing as some of the most persistent and ever-present wrongs of deepfakes even if they do not pose substantial political or personal harms.

reasons. If artifactual deliberation tracks the practices of epistemic agents, we might then expect practices of deliberation to actually worsen the harms produced by nonconsensual pornography.

Given these considerations, when we adopt artifactual deliberation as a model for understanding the social verification of deepfakes, the result is neither broad skepticism toward recordings nor is it a callous rejection of any and all novel harms posed by synthetic media. Rather, artifactual deliberation deflates certain epistemic harms while also identifying situations where there are limited resources for verifying manipulated audio-visual media or where harms cannot be mitigated by our epistemic practices. As a methodological intervention into empirical assessments of deepfakes, this position suggests that future research on deepfake detection should consider the effects of socially structured, media-saturated environments on human assessments in order to avoid perpetuating the AI arms race narrative and its associated problems. Additionally, more focus should be placed on the epistemic and ethical harms which lie beyond the framing of the epistemic arms race. Finally, I have suggested that the prospect of artifactual deliberation shifts social epistemology closer to positions offered in communication and media studies. Hopefully, with novel sites for cross-disciplinary collaboration we can better frame the current epistemic threats of political disinformation while also remaining sensitive to those less visible harms which matter most.

# CHAPTER 6

## CONCLUSION:

## TECHNOLOGY, NORMATIVITY, AND THE PRAGMATIST ENLIGHTENMENT

> The space of reasons has been constrained before—by priests, kings, oligarchs, and family elders who would gladly substitute their moral and political judgments for ours, and by bureaucrats who endlessly invent analog means of rendering their own judgments opaque. But at the heart of the modern Enlightenment lies the eighteenth-century philosopher Immanuel Kant's urgent call: sapere aude! He asked us to dare to think for ourselves, together. [...] Today we risk losing that inheritance to the algorithms engraved in our "helpful" AI mirrors, which we are told can progress civilization by performing all the critical operations of moral and political life without having to think about them.
>
> — Shannon Vallor, *The AI Mirror* (2024)

In recent decades, several key pragmatist philosophers have called for a new Enlightenment divorced from epistemological foundationalism, one which reconstrues the task of normative autonomy as a social achievement to be enacted through democratic society rather than through an individual subject's representational access to reality. In his last published work, Richard Rorty (2021) reiterates the overarching theme of his pragmatism as a species of anti-authoritarianism which extends the Enlightenment's criticism of theological authority to the authority supposedly imposed upon us by non-human reality. In Rorty's view, just as the Enlightenment developed socio-political arguments against the divine right of kings and the theological supremacy of the Church, so too did it create its own phantoms in the form of vacuous appeals to an ultimate truth and reality. As Robert Brandom (2021) elaborates the thesis, Rorty's last project diagnoses a central tension in Enlightenment thought. On one hand, Enlightenment philosophers sought to undermine non-rational appeals to authority in political life. On the other hand, they made liberal use of unjustified justifiers in their theoretical endeavors (2021, xxiii). Drawing out this tension, Rorty's pragmatism calls for a global anti-authoritarianism which criticizes the use of unjustified justifiers in any and all domains. This amounts to a radical kind of social arrangement where every

rational individual must be prepared to give reasons for their beliefs and any individual who refuses this requirement for reason-giving expresses an unjustified authority over their peers.

After decades of disagreements with Rorty concerning the course and scope of pragmatism, Hilary Putnam nonetheless equally champions a "pragmatist Enlightenment" (2004, 96). As with Rorty, Putnam thinks the Enlightenment's appeal to foundationalist epistemology devolves into either rationalist dogmatism or empiricist fabulation (100). In Putnam's positive formulation of the pragmatist Enlightenment, the classical pragmatists, especially Dewey, deny the notion of a self-sufficient epistemological or moral subject. Instead, Dewey maintained that the "reflective transcendence" of the Enlightenment is to be achieved through democratic arrangements of society, through better, more inclusive, and more intelligent practices (101). As with Rorty, Putnam's challenge to the Enlightenment amounts to a criticism of methodological individualism. The goal of democracy is to engender a self-correcting process where no participant goes unheard and no question goes unanswered. Ironically, the efforts of modern epistemology implicitly undermined this process by demarcating a class of answers which could never be questioned, and it set into motion a quest for certainty that diverted philosophical attention away from the social practices underlying knowledge production.

Robert Brandom (2011) draws from these two accounts of the pragmatist Enlightenment to argue that pragmatism's historical weight comes from its intermixing of rationalism and romanticism, its combination of the critical Kantian thesis that reason is the source of human autonomy with the creative impulse that each individual has a unique inner light. In combining these two threads, Brandom's construal of the pragmatist Enlightenment involves the shift from an understanding of reason as an individual faculty of the mind to an understanding of reason as a social practice:

> The European Enlightenment had seen the "natural light of reason" as universal in the sense of *shared*, or *common*, so that what one disinterested, selfless scientist could add as a brick to the edifice of knowledge, another could in principle do as well. By contrast the pragmatists, looking at the division of labor in what had become modern industrial economy, saw the enterprise of reason as *social* in a more genuine, articulated, *ecological* sense, in which the contributions of individuals are not interchangeable or fungible, in which each has potentially a unique contribution to make to the common enterprise, which requires many different sorts of skills, responses, ideas, and assessments, all collectively serving as the environment in which each adapts and evolves. (Brandom 2011, 41)

Though Brandom's appropriation of the pragmatist Enlightenment creatively rereads Dewey's experimentalism and Rorty's ironist conversationalism in uncharacteristically rationalist terms, there is no denying the power of his narrative. What Brandom is after is a historical backdrop for the space of reasons, one which demonstrates the social importance of his choice of philosophical method.

And yet I have been claiming that something is missing in the positions Rorty and Brandom occupy, that their conception of the space of reasons is too narrowly construed to perform the kind of social and historical work for which it has been developed. My contention is that our space of reasons in the twenty-first century looks very different from that of James and Dewey or Kant and Voltaire. Certainly, the rules of reasoning are the same. There are roughly the same norms of inference between languages and across cultures for us to reconstruct history in Rorty's (1979) terms as a singular, grand "conversation of mankind." But the difference is that the objects of the game of giving and asking for reasons have changed despite the functional structure remaining stable. These kinds of changes in objects are permitted in any set of practices. It is just like how chess can now be played virtually and Bach's fugues are now often performed with pianos rather than harpsichords, yet we do not regard virtual chess as any different from wooden chess nor do we claim that contemporary pianists never actually play Bach.

However, these minute differences matter. The two-dimensional presentation of the virtual chess board may impact the speed, accuracy, and confidence of the player's moves and the sustain or sostenuto pedals may be employed by the pianist to let a note linger longer than once possible on the 18$^{th}$ century clavier. In the first case, changing objects at play within the practice changes the effectiveness of the performance. In the second case, changing the objects alters the available performances altogether.

Artifactual pragmatism holds that these two kinds of modifications equally apply to the game of giving and asking for reasons. New technologies are readily integrated into the space of reasons when rational interlocutors accept these technologies as providing reliable and unbiased assessments toward a specific class of properties, and these technologies subsequently alter the common moves made in practices of reason-giving. The artifactual pragmatist still sees a common "conversation" of mankind, a singular space of reasons, but they further investigate how modifications in artifactual arrangements produce modifications to common and acceptable acts of justification. In this manner, artifactual pragmatism can tell a progressive story about how

appeals to public artifacts adjudicate epistemic disagreements that were previously unresolvable and it can tell a regressive story about how some new artifacts undercut the effectiveness of prior modes of justification.

Though Brandom (2011) observes that the pragmatist Enlightenment resulted from the rapid advance of technology in the latter 19th century, the impacts of technology never seem to creep into the practices of reasoning themselves. Brandom writes that, for the classical pragmatists, "technology *embodies* understanding" (2011, 38), but he, himself, cannot accept this premise. Unlike Dewey and Rorty, he is too suspicious of instrumentality and too taken by the technophobia common among post-Heideggerian philosophers. Rorty is more willing to grant technology a positive influence and a more central role in his philosophy, but he equally relegates non-linguistic artifacts to a position outside of the space of reasons. For Brandom, technology must be separated from reason to maintain the purity of reason. For Rorty, technology must be separated to make room for feeling, sentiment, and skillful coping. The problem for Brandom is that practices are never pure. The problem for Rorty is that technologies are never *just* devices for skillful coping.

Earlier, I foregrounded Barry Allen's (2000, 2004) oft-neglected criticism that Rorty holds a propositional bias throughout his work. Unlike Rorty's opponents who accuse him of relativism or skepticism, Allen's criticism strategically indicts Rorty on his own terms, implying that Rorty's propositional bias is a holdover of the same representationalist epistemology he so deplores. With a view towards Rorty's anti-authoritarian modification to the Enlightenment, Allen's criticism is more forcefully reconstructed in normative terms. He is claiming that Rorty's pragmatism is still authoritarian, that Rorty's propositional bias is an unjustified justifier upon which his project rests.

While I do believe that warrant for Rorty's propositional bias can be provided, I have been showing that it is a fruitful exercise to explore the benefits that come with rejecting his propositional bias while still preserving the rationalist framework of the space of reasons. I have shown that this combination of commitments enables artifacts to serve as reasons in a manner that can be readily imported to solve concrete problems where the primary content under investigation is multimodal or audio-visual in nature.

Moreover, my account revises the standard picture of the pragmatist Enlightenment by reframing how technologies are embedded within those very practices that constrain unjustified authority. Here, one can point to several historical instances where non-linguistic artifacts justified what would otherwise have perhaps remained unjustifiable. Photographs of the scars crisscrossing

the backs of enslaved Black men challenge the myth of benevolent slavery perpetuated by whites in the American south and film recordings shown at Nuremberg of the mass graves at Bergen-Belsen still provide some of the most extensive evidence of the indescribable horrors of Nazi concentration camps. The scenes captured by these recordings are alarming and disturbing. Their immediate, visceral effect explains why one might initially regard them as appealing to pathos rather than logos, of affectively stirring up one's interlocutor's moral sentiments instead of justifying their beliefs. But this is a false dichotomy. Recordings can act as objects of justification, as reasons, just as they stir up feelings of sorrow, disgust, and moral outrage.

In Chapter 3, I showed that the pragmatist's propositional bias has continuously been utilized to deny the evidential function of recordings. At times, the move has been to treat recordings as expressing implicit propositions; this is the more innocuous position. But at other moments, the bias seems to entirely dismiss or disregard the function that these recordings have played in documenting and ultimately denouncing authoritarianism, and this dismissal is primarily motivated in order to preserve some favored theoretical premises. My worry is that these theoretical premises have harmed contemporary pragmatism. Ultimately, the propositional bias obscures how practices of justification actually operate, preserves an artificial divide between philosophical and non-philosophical approaches to communication which inhibits collaboration, and at its worst, trivializes non-linguistic evidence by comparing the utilization of such evidence to meaningless babblings and even enactments of physical violence.

Though I am impressed by the vision of Rorty's pragmatist Enlightenment and wholeheartedly endorse his anti-authoritarianism, I am also suspicious at where the line between authoritarianism and anti-authoritarianism has been drawn. In accordance with Rorty's own project, I have stressed that one need not give up anti-representationalism in epistemology when addressing the normative and epistemic role of artifacts. If anything, a pragmatist account of artifacts reaffirms and makes explicit the Enlightenment's underlying dependency on mechanical and instrumental knowledge. It tracks how reliable instruments have become increasingly entangled with the spoken and written word in everyday practices of justification.

In the twenty-first century, our practices are more deeply entangled with technology than ever before, and it is an ever-greater imperative that we interrogate the relationship of technology and practice to avoid trading our hard-won normative autonomy for a set of more effective tools. Pragmatism is especially situated to enact this investigation, as it begins with actual practices in

order to extrapolate theoretical and conceptual insights rather than first postulating the abstract insights in order to retrofit practices. In this manner, pragmatism is opposed to the method of moral inquiry which utilizes otherworldly thought experiments to derive "self-evident" moral intuitions. As Elizabeth Anderson explains, these thought experiments are cases of "simulated deliberation" which remove the actual stakes and consequences of acting (2021, 85).

In the philosophy of technology, the outcomes of this abstract method of moral theorizing pose a disruption to normative autonomy. Philosophers of technology are especially prone to using simulated deliberation in order to assess sci-fi-like examples for their catastrophic risks. On the reverse side, sociologically-informed perspectives into technology may dismiss the need for normativity. For example, the "empirical turn" into social effects of technology, which emerged in the 1980s and 1990s through interdisciplinary fields such as STS, often operated on a purely descriptive level, one which avoided prescriptive language altogether (Verbeek 2011, 161–162; Vallor 2016, 31–32). To avoid both abstractionism and descriptivism, pragmatism asserts that ethical analyses into technology may readily utilize empirical approaches in order to provide the material for normative analyses. As Shannon Vallor puts it, "the empirical turn can *feed* practical wisdom by opening up ever more avenues for gathering vital contextual information about new and emerging technosocial realities" (2016, 33).

It is this normative groundedness which protects pragmatism from readily aligning itself with either naïve accounts of progress or with pessimistic narratives of impending disaster. For this reason, David S. Watson, Jakob Mökander, and Luciano Floridi, defend a *sociotechnical pragmatism* which avoids both deterministic narratives of rapid technological advancement and overly critical evaluations of technology which threaten their own ethical and epistemic basis (Watson et al. 2025). This account reaffirms that pragmatism is neither utopian nor pessimistic with respect to technology. Rather, pragmatism brackets out present problems within practices and then defends the possibility of ameliorating these problems through extensions of experimentation into ethical life.

Finally, pragmatism combines the romantic and the rational in ways that extend beyond social justification. Technologies not only enter into the space of reasons when we use artifacts to justify, but they also constitute the space of practice, where we interact with artifacts in ways which are primarily affective or aesthetic. A pragmatist ethics of technology thus combines the rational "context of justification" with the romantic "context of discovery" (Keurlartz et al. 2004, 19). Or

in other words, it accounts for the informal, creative impulses which underlie and intermix with the formal giving and asking for reasons within democratic societies.

By situating itself between abstractionism and descriptivism, optimism and pessimism, rationalism and romanticism, philosophical pragmatism offers one of the most robust normative frameworks for the philosophy of technology. Mediating these contrasts, this dissertation is one brief attempt to study the relationship between technology and practice through a pragmatist lens, paying close attention to different manifestations of the peculiar practice of reason-giving. In keeping with the pragmatist spirit, I have argued that this practice looks very different from prior iterations by foregrounding current contexts of justification which challenge the notion that the space of reasons is a wholly linguistic affair. I have therefore treated the space of reasons as a technology to be analyzed, as a mechanism where repairs must be made and where new parts are haphazardly added. However, more work must be done in order to illuminate the role of technology within the pragmatist Enlightenment, and at present, my own work has not yet fully surveyed the relevant domains of empirical research into generative AI, nor has it fully invested itself in the theoretical import of a non-rational space of practice.

Nonetheless, I hope to have added a new part to the underlying mechanism of the space of reasons which opens it up to further research. My addition, artifactual pragmatism, emphasizes the material aspects in that same process that Sellars, Rorty, and Brandom have identified as social and it further attacks the notion that reasoning is a black box enclosed within the head. It holds that meaning extends across our common artifacts and leverages this insight to rethink the boundaries of where practices of justification begin and end. Finally, artifactual pragmatism extends the pragmatist's commitment to the Enlightenment project, for it sees that anti-authoritarianism is dependent upon access to and distributions of non-linguistic social technologies. Thus, we should not downplay the technological within our accounts of the rational nor should we treat technology and reason as opposing factions. Technology does not threaten our rational autonomy. Rather, by properly accounting for technology in the space of reasons, we can show how anti-authoritarian technologies are possible.

# BIBLIOGRAPHY


Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. 2019. *The State of Deepfakes: Landscape, Threats, and Impact*. DeepTrace.

Alexander, Thomas. 2013. *The Human Eros: Eco-Ontology and the Aesthetics of Existence*. Fordham University Press. https://doi.org/10.5422/fordham/9780823251209.001.0001.

——. 2014. "Linguistic Pragmatism and Cultural Naturalism: Noncognitive Experience, Culture, and the Human Eros." *European Journal of Pragmatism and American Philosophy* VI (2): 2. https://doi.org/10.4000/ejpap.299.

Allen, Barry. 2000. "What Was Epistemology?" In Brandom 2000b.

——. 2004. *Knowledge and Civilization*. Westview Press.

——. 2008. *Artifice and Design: Art and Technology in Human Experience*. Cornell University Press.

——. 2013. "The Use of Useless Knowledge: Bergson against the Pragmatists." *Canadian Journal of Philosophy* 43 (1): 37–59.

——. 2023. *Living in Time: The Philosophy of Henri Bergson*. Oxford University Press.

Allyn, Bobby. 2022. "Deepfake Video of Zelenskyy Could Be 'tip of the Iceberg' in Info War, Experts Warn." Technology. *NPR*, March 16. https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia.

Alvarado, Ramón. 2022. "Should We Replace Radiologists with Deep Learning? Pigeons, Error and Trust in Medical AI." *Bioethics* 36 (2): 121–33. https://doi.org/10.1111/bioe.12959.

——. 2023. "AI as an Epistemic Technology." *Science & Engineering Ethics* 29 (5): 1–30. https://doi.org/10.1007/s11948-023-00451-3.

Anderson, Elizabeth. 2006. "The Epistemology of Democracy." *Episteme: A Journal of Social Epistemology* 3 (1): 8–22. https://doi.org/10.1353/epi.0.0000.

——. 2021. "How to Be a Pragmatist." In *The Routledge Handbook of Practical Reason*, edited by Ruth Chang and Kurt Sylvan. Routledge.

Arendt, Hannah. 1958. *The Human Condition*. University of Chicago Press.

Aristotle. 1941. *Metaphysics*. In *The Basic Works of Aristotle*, edited by Richard McKeon. Random House

Arora, Chirag. 2024. “Proxy Assertions and Agency: The Case of Machine-Assertions.” *Philosophy & Technology* 37 (1): 15. https://doi.org/10.1007/s13347-024-00703-5.

Armstrong, David M. 1973. *Belief, Truth and Knowledge*. Cambridge University Press.

——. 1993. “A World of States of Affairs.” *Philosophical Perspectives* 7: 429–40. https://doi.org/10.2307/2214133.

Atencia-Linares, Paloma. 2012. “Fiction, Nonfiction, and Deceptive Photographic Representation.” *The Journal of Aesthetics and Art Criticism* 70 (1): 19–30.

Atencia-Linares, Paloma, and Marc Artiga. 2022. “Deepfakes, Shallow Epistemic Graves: On the Epistemic Robustness of Photography and Videos in the Era of Deepfakes.” *Synthese* 200 (6): 1–22. https://doi.org/10.1007/s11229-022-04003-3.

Austin. J. L. 1964. *Sense and Sensibilia*. Oxford University Press.

Ayers, Michael. 2004. “Sense Experience, Concepts and Content, Objections to Davidson and McDowell.” In *Perception and Reality - From Descartes to the Present*, edited by Ralph Schumacher. Mentis.

Bagwala, Abbas. 2026. “When Does My AI Reason? Sapience and the Inferential Grounding Condition,” Unpublished manuscript, April 21, 2026.

Baird, Davis. 2004. *Thing Knowledge: A Philosophy of Scientific Instruments*. University of California Press.

Bender, Emily M., and Alexander Koller. 2020. “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜.” *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA), FAccT ’21, March 1, 610–23. https://doi.org/10.1145/3442188.3445922.

Bergson, Henri. (1911) 1998. *Creative Evolution*. Translated by Arthur Mitchell. Dover Publications.

Bernstein, Richard J. 2010. *The Pragmatic Turn*. Polity Press.

Birdsell, David S., and Leo Groarke. 2007. “Outlines of a Theory of Visual Argument.” *Argumentation and Advocacy* 43 (3–4): 103–13. https://doi.org/10.1080/00028533.2007.11821666.

Birrer, Alena, and Natascha Just. 2024. “What We Know and Don’t Know about Deepfakes: An Investigation into the State of the Research and Regulatory Landscape.” *New Media & Society*, May 22, 1–20. https://doi.org/10.1177/14614448241253138.

Blackburn, Simon. 2019. “Wittgenstein and Brandom: Affinities and Divergences.” *Disputatio* 8 (9): 111–131.

Blair, J. Anthony. 1996. “The Possibility and Actuality of Visual Arguments.” *Argumentation and Advocacy* 33 (1): 23.

——. 2015. “Probative Norms for Multimodal Visual Arguments.” *Argumentation* 29 (2): 217–33. https://doi.org/10.1007/s10503-014-9333-3.

Block, Ned. 1981. “Psychologism and Behaviorism.” *The Philosophical Review* 90 (1): 5–43. https://doi.org/10.2307/2184371.

Brandom, Robert. B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

——. 1997. “Study Guide” In *Empiricism and the Philosophy of Mind*, edited by Robert B. Brandom. Harvard University Press.

——.1998. “Insights and Blindspots of Reliabilism.” In Brandom 2000a.

——. 2000a. *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press.

——. ed. 2000b. *Rorty and His Critics*. Blackwell.

——. 2000c. “Vocabularies of Pragmatism” In Brandom 2000c.

——. 2008. *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford University Press.

——. 2009a. How Analytic Philosophy Has Failed Cognitive Science. In Brandom 2009b.

——. 2009b. *Reason in Philosophy: Animating Ideas*. Harvard University Press.

——. 2011. *Perspectives on Pragmatism: Classical, Recent, and Contemporary*. Harvard University Press.

——. 2019. “Some Strands of Wittgenstein’s Normative Pragmatism, and Some Strains of His Semantic Nihilism.” *Disputatio* 8 (9): 9. https://doi.org/10.5281/zenodo.2631339.

——. 2021. Foreword to *Pragmatism as Anti-Authoritarianism* by Richard Rorty. Edited by Eduardo Mendieta. Harvard University Press.

Brandom, Russell. 2019. "Deepfake Propaganda Is Not a Real Problem." *The Verge*, March 5. https://www.theverge.com/2019/3/5/18251736/deepfake-propaganda-misinformation-troll-video-hoax.

Bray, Sergi D, Shane D Johnson, and Bennett Kleinberg. 2023. "Testing Human Ability to Detect 'Deepfake' Images of Human Faces." *Journal of Cybersecurity* 9 (1): tyad011. https://doi.org/10.1093/cybsec/tyad011.

Breland, Ali. 2019. "The Bizarre and Terrifying Case of the 'Deepfake' Video That Helped Bring an African Nation to the Brink." *Mother Jones*, March 15. https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/.

Brewer, Bill. 2011. *Perception and Its Objects*. Oxford University Press.

Brown, Julia E. H., and Jodi Halpern. 2021. "AI Chatbots Cannot Replace Human Interactions in the Pursuit of More Inclusive Mental Healthcare." *SSM - Mental Health* 1 (December): 100017. https://doi.org/10.1016/j.ssmmh.2021.100017.

Bucher, Taina, and Anne Helmond. 2018. "The Affordances of Social Media Platforms." In *The SAGE Handbook of Social Media*, by Jean Burgess, Alice Marwick, and Thomas Poell. SAGE Publications.

Burke, Thomas. 1994. *Dewey's New Logic: A Reply to Russell*. University of Chicago Press.

Burkell, Jacquelyn, and Chandell Gosse. 2019. "Nothing New Here: Emphasizing the Social and Cultural Context of Deepfakes." *First Monday* 24 (12). https://doi.org/10.5210/fm.v24i12.10287.

Caplan, Asher Brandon. 2025. "Technology and Social Practice: Reassessing the Language-as-Tool Metaphors in Dewey and Wittgenstein." *Contemporary Pragmatism* 1 (aop): 1–25. https://doi.org/10.1163/18758185-bja10118.

Cavedon-Taylor, Dan. 2013. "Photographically Based Knowledge." *Episteme* 10 (3): 283–97. https://doi.org/10.1017/epi.2013.21.

——. 2024. "Deepfakes: A Survey and Introduction to the Topical Collection." *Synthese* 204 (1): 1–19. https://doi.org/10.1007/s11229-024-04634-8.

Chalmers, David J. 2022. *Reality+: Virtual Worlds and the Problems of Philosophy*. W. W. Norton & Company.

——. 2023. “Could a Large Language Model Be Conscious?” *Boston Review*, August 9. https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/.

Chandrasekar, Aisvarya, and Klaudia Jaźwińska. 2025. “Why AI Models Are Bad at Verifying Photos.” *Columbia Journalism Review*. August 26. https://www.cjr.org/tow_center/why-ai-models-are-bad-at-verifying-photos.php.

Chang, Hasok. 2007. *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press.

——. 2022. *Realism for Realistic People: A New Pragmatist Philosophy of Science*. Cambridge University Press.

Chesney, Robert and Danielle Citron. 2019. “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security.” *California Law Review* 107 (6): 1753.

Churchland, Patricia S. 2013. Foreword to *Word and Object.* New ed, by Wilard van Orman Quine. Harvard University Press.

Coeckelbergh, Mark. 2017. *Using Words and Things: Language and Philosophy of Technology*. New York: Routledge.

Cole, Samantha. 2017. “AI-Assisted Fake Porn Is Here and We’re All Fucked.” *Vice*, December 11. https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn.

Colombatto, Clara, and Stephen M Fleming. 2024. “Folk Psychological Attributions of Consciousness to Large Language Models.” *Neuroscience of Consciousness* 2024 (1): niae013. https://doi.org/10.1093/nc/niae013.

Cuevas-Badallo, Ana. 2021. “A Pragmatist Explanation of Technical Capabilities in Nonhuman Animals.” *European Journal of Pragmatism and American Philosophy* XIII (1): 1. https://doi.org/10.4000/ejpap.2324.

Cunningham, J. J. 2018. “Are Perceptual Reasons the Objects of Perception?” In *In the Light of Experience: New Essays on Perception and Reasons*, edited by Johan Gersel, Rasmus Thybo Jensen, Morten S. Thaning, and Søren Overgaard. Oxford University Press.

Davidson, Donald. 1973. “On the Very Idea of a Conceptual Scheme.” In Davidson 1984.

——. 1978. “What Metaphors Mean” in Davidson 1984.

——. 1983. “A Coherence Theory of Truth and Knowledge.” In Davidson 1984.

——. 1984. *Inquiries into Truth and Interpretation*. Oxford University Press.

——. 1986. “A Nice Derangement of Epitaphs.” In Davidson 2005.

——. 1987. “Knowing One’s Own Mind.” In Davidson 2001.

——. 1990a. “Epistemology Externalized” In Davidson 2001.

——. 1990b. “Meaning, Truth, and Evidence.” In Davidson 2005.

——. 1990c. “Turing’s Test.” In Davidson 2004.

——. 1992. “The Third Man.” In Davidson 2005.

——. 1997a. “Seeing Through Language.” In Davidson 2005.

——. 1997b. “The Emergence of Thought” In Davidson 2001.

——. 1998. “The Irreducibility of the Concept of the Self” In Davidson 2001.

——. 2000. “Truth Rehabilitated.” In Davidson 2005.

——. 2001. *Subjective, Intersubjective, Objective*. Oxford University Press.

——. 2004. *Problems of Rationality*. Oxford University Press.

——. 2005. *Truth, Language and History*. Oxford University Press.

Dennett, Daniel C. 1987. *The Intentional Stance*. MIT Press.

——. 1997. “When HAL kills, who's to blame?” In *HAL's Legacy: 2001's Computer as Dream and Reality,* edited by David G. Stork. MIT Press.

deVries, Willem A. 2005. *Wilfrid Sellars*. McGill-Queen’s University Press.

Dewey, John. 1969–1999. *The Collected Works of John Dewey (Early Works, Middle Works, and Late Works)*, edited by Jo Ann Boydston. Carbondale, IL: Southern Illinois University Press.

——. 1896. “The Reflex Arc Concept in Psychology.” In Dewey EW 5.

——. 1916a. *Democracy and Education.* In Dewey MW 9.

——. 1916b. “Logical Objects.” In Dewey MW 10.

——. 1922. *Human Nature and Conduct.* In Dewey MW 14.

——. 1925. *Experience and Nature*. In Dewey LW 1.

——. 1938. *Logic: The Theory of Inquiry*. In Dewey LW 12.

——. 1949. “John Dewey to John D. Graves (11008):” In *The Correspondence of John Dewey, Vol 3: 1940-1953,* electronic ed, edited by Larry Hickman. Center for Dewey Studies, Southern Illinois University.

Diderot, Denis. 2003. “Man.” In *The Encyclopedia of Diderot & d'Alembert Collaborative Translation Project.* Translated by Nelly S. Hoyt and Thomas Cassirer. Ann Arbor: Michigan Publishing, University of Michigan Library, http://hdl.handle.net/2027/spo.did2222.0000.160. Originally published as "Homme," *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers,* 8:256–257 (Paris, 1765).

Dieleman, Susan. 2025. “Norms and Novelty in the Space of Reasons.” *European Journal of Pragmatism and American Philosophy* 17 (2). https://doi.org/10.4000/15d8j.

Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, et al. 2021. “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.” *arXiv*, June 3. https://doi.org/10.48550/arXiv.2010.11929.

Dretske, Fred. 1989. “Reasons and Causes.” *Philosophical Perspectives* 3: 1–15. https://doi.org/10.2307/2214261.

Dreyfus, Hubert L. 2005. “Overcoming the Myth of the Mental: How Philosophers Can Profit from the Phenomenology of Everyday Expertise.” *Proceedings and Addresses of the American Philosophical Association* 79 (2): 47–65.

Durán, Juan M., and Nico Formanek. 2018. “Grounds for Trust: Essential Epistemic Opacity and Computational Reliabilism.” *Minds and Machines* 28 (4): 645–66. https://doi.org/10.1007/s11023-018-9481-6.

Ebbs, Gary. 2025. “Content Externalism and Skepticism.” In *The Stanford Encyclopedia of Philosophy*, Spring 2025, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2025/entries/skepticism-content-externalism/.

Engler, Alex. 2019. “Fighting Deepfakes When Detection Fails.” *Brookings Institution*, November 14. https://www.brookings.edu/articles/fighting-deepfakes-when-detection-fails/.

Fallis, Don. 2021. “The Epistemic Threat of Deepfakes.” *Philosophy & Technology*, 30: 623–643. https://doi.org/10.1007/s13347-020-00419-2.

Fisher, Sarah A. 2024. “Large Language Models and Their Big Bullshit Potential.” *Ethics and Information Technology* 26 (4): 67. https://doi.org/10.1007/s10676-024-09802-5.

Fisher, Sarah A., Jeffrey W. Howard, and Beatriz Kira. 2024. “Moderating Synthetic Content: The Challenge of Generative Ai.” *Philosophy and Technology* 37 (4): 1–20. https://doi.org/10.1007/s13347-024-00818-9.

Flattery, Tobias, and Christian B. Miller. 2024. “‘Deepfakes and Dishonesty.’” *Philosophy & Technology* 37 (4): 120. https://doi.org/10.1007/s13347-024-00812-1.

Floridi, Luciano. 2011. *The Philosophy of Information*. Oxford University Press.

Franks, Mary Anne, and Ari Waldman. 2019. “Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions.” *Maryland Law Review* 78 (January): 892.

Frápolli, María José. 2023. *The Priority of Propositions: A Pragmatist Philosophy of Logic*. Springer Nature.

Frápoli, María José, and Kurt Wischin. 2019. “From Conceptual Content in Big Apes and AI, to the Classical Principle of Explosion: An Interview with Robert B. Brandom.” *Disputatio* 8 (9): 9. https://doi.org/10.5281/zenodo.2642444.

Fesmire, Steven. 2014. *Dewey*. Routledge.

Freiman, Ori. 2023. “Analysis of Beliefs Acquired from a Conversational AI: Instruments-Based Beliefs, Testimony-Based Beliefs, and Technology-Based Beliefs.” *Episteme*, March 6, 1–17. https://doi.org/10.1017/epi.2023.12.

Freiman, Ori, and Boaz Miller. 2020. “Can Artificial Entities Assert?” In *The Oxford Handbook of Assertion*, edited by Sanford Goldberg. Oxford University Press.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

Frosolini, Andrea, Paolo Gennaro, Flavia Cascino, and Guido Gabriele. 2023. “In Reference to ‘Role of Chat GPT in Public Health’, to Highlight the AI’s Incorrect Reference Generation.” *Annals of Biomedical Engineering* 51 (10): 2120–22. https://doi.org/10.1007/s10439-023-03248-4.

Gao, Catherine A., Frederick M. Howard, Nikolay S. Markov, et al. 2023. “Comparing Scientific Abstracts Generated by ChatGPT to Real Abstracts with Detectors and Blinded Human Reviewers.” *Npj Digital Medicine* 6 (1): 75. https://doi.org/10.1038/s41746-023-00819-6.

Gao, Yunfan, Yun Xiong, Xinyu Gao, et al. 2024. “Retrieval-Augmented Generation for Large Language Models: A Survey.” arXiv:2312.10997. *arXiv*, March 27. https://doi.org/10.48550/arXiv.2312.10997.

Godden, David. 2017. “On the Norms of Visual Argument: A Case for Normative Non-Revisionism.” *Argumentation* 31 (2): 395–431. https://doi.org/10.1007/s10503-016-9411-9.

——. 2022. “Logics for ‘Non-Logical’ Argumentation: A ‘Neo-Logicist’ Defense of the Primacy of the ‘Logical’ Mode of Argument in Gilbert’s Multi-Modal Theory of Argumentation.” *Informal Logic* 42 (3): 521–62. https://doi.org/10.22329/il.v42i3.7499.

Godler, Yigal, Zvi Reich, and Boaz Miller. 2020. “Social Epistemology as a New Paradigm for Journalism and Media Studies.” *New Media & Society* 22 (2): 213–29. https://doi.org/10.1177/1461444819856922.

Goh, Dion Hoe-Lian, Jonathan Pan, and Chei Sian Lee. 2024. “Humans Versus Machines: A Deepfake Detection Faceoff.” *Proceedings of the Association for Information Science and Technology* 61 (1): 917–19. https://doi.org/10.1002/pra2.1139.

Goldberg, Sanford C. 2012. “Epistemic Extendedness, Testimony, and the Epistemology of Instrument-Based Belief.” *Philosophical Explorations* 15 (2): 181–97. https://doi.org/10.1080/13869795.2012.670719.

——. 2020. “Epistemically Engineered Environments.” *Synthese* 197 (7): 2783–802.

Goldman, Alvin I. 1976. “Discrimination and Perceptual Knowledge.” *The Journal of Philosophy* 73 (20): 771–91. https://doi.org/10.2307/2025679.

——. 1979. “What Is Justified Belief?” In *Justification and Knowledge*, edited by George Pappas. D. Reidel.

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, et al. 2014. “Generative Adversarial Networks.” *arXiv*, June 10. https://doi.org/10.48550/arXiv.1406.2661.

Gosse, Chandell, and Jacquelyn Burkell. 2020. “Politics and Porn: How News Media Characterizes Problems Presented by Deepfakes.” *Critical Studies in Media Communication* 37 (5): 497–511. https://doi.org/10.1080/15295036.2020.1832697.

Groarke, Leo. 1996. “Logic, Art and Argument.” *Informal Logic* 18 (2): 105–129. https://doi.org/10.22329/il.v18i2.2376.

——. 2015. “Going Multimodal: What Is a Mode of Arguing and Why Does It Matter?” *Argumentation* 29 (2): 133–55. https://doi.org/10.1007/s10503-014-9336-0.

Groh, Matthew, Ziv Epstein, Chaz Firestone, and Rosalind Picard. 2022. “Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds.” *Proceedings of the National Academy of Sciences of the United States of America* 119 (1): e2110013119. https://doi.org/10.1073/pnas.2110013119.

Gross, Alan G. 2009. “Toward a Theory of Verbal–Visual Interaction: The Example of Lavoisier.” *Rhetoric Society Quarterly* 39 (2): 147–69. https://doi.org/10.1080/02773940902766755.

Grzankowski, Alex. 2015. “Pictures Have Propositional Content.” *Review of Philosophy and Psychology* 6 (1): 151–63. https://doi.org/10.1007/s13164-014-0217-0.

Gubelmann, Reto. 2023. “A Loosely Wittgensteinian Conception of the Linguistic Understanding of Large Language Models like BERT, GPT-3, and ChatGPT.” *Grazer Philosophische Studien* 99 (4): 485–523. https://doi.org/10.1163/18756735-00000182.

Haack, Susan. 2004. “Pragmatism, Old and New.” *Contemporary Pragmatism* 1 (1): 3–41. https://doi.org/10.1163/18758185-90000126.

Harnad, Stevan. 1990. “The Symbol Grounding Problem.” *Physica D: Nonlinear Phenomena* 42 (1): 335–46. https://doi.org/10.1016/0167-2789(90)90087-6.

Havlík, Vladimír. 2025. “Meaning and Understanding in Large Language Models.” *Synthese* 205 (1): 9. https://doi.org/10.1007/s11229-024-04878-4.

Habgood-Coote, Joshua. 2023. “Deepfakes and the Epistemic Apocalypse.” *Synthese* 201 (3): 103. https://doi.org/10.1007/s11229-023-04097-3.

Hameleers, Michael, Thomas E. Powell, Toni G.L.A. Van Der Meer, and Lieke Bos. 2020. “A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media.” *Political Communication* 37 (2): 281–301. https://doi.org/10.1080/10584609.2019.1674979.

Harris, Keith Raymond. 2021. “Video on Demand: What Deepfakes Do and How They Harm.” *Synthese* 199 (5–6): 13373–91. https://doi.org/10.1007/s11229-021-03379-y.

——. 2024. “AI or Your Lying Eyes: Some Shortcomings of Artificially Intelligent Deepfake Detectors.” *Philosophy & Technology* 37 (1): 7. https://doi.org/10.1007/s13347-024-00700-8.

Heinrichs, Bert, and Sebastian Knell. 2021. “Aliens in the Space of Reasons? On the Interaction Between Humans and Artificial Intelligent Agents.” *Philosophy & Technology* 34 (4): 1569–80. https://doi.org/10.1007/s13347-021-00475-2.

Hickman, Larry A. 1990. *John Dewey’s Pragmatic Technology*. Indiana University Press.

——. 2001. *Philosophical Tools for Technological Culture: Putting Pragmatism to Work*. Indiana University Press.

——. 2013. “Technological Pragmatism.” In *A Companion to the Philosophy of Technology*, edited by Jan Kyrre Berg Olsen Friis, Stig Andur Pedersen, and Vincent F. Hendricks. Blackwell

Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. “ChatGPT Is Bullshit.” *Ethics and Information Technology* 26 (2): 38. https://doi.org/10.1007/s10676-024-09775-5.

Hildebrand, David L. 2003. *Beyond Realism and Antirealism: John Dewey and the Neopragmatists*. Vanderbilt University Press.

Horkheimer, Max. 1947. *Eclipse of Reason*. Oxford University Press.

Huang, Lei, Weijiang Yu, Weitao Ma, et al. 2025. “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions.” *ACM Transactions on Information Systems* 43 (2): 1–55. https://doi.org/10.1145/3703155.

Huetter-Almerigi, Yvonne. 2020. “Two Forms of Realism.” *European Journal of Pragmatism and American Philosophy* XII (1). https://doi.org/10.4000/ejpap.1868.

Hurley, Susan. 2003. “Animal Action in the Space of Reasons.” *Mind and Language* 18 (3): 231–57. https://doi.org/10.1111/1468-0017.00223.

Hyska, Megan. 2025. “The Politics of Past and Future: Synthetic Media, Showing, and Telling.” *Philosophical Studies* 182 (1): 137–58. https://doi.org/10.1007/s11098-023-02062-x.

Johnson, Mark. 2008. *The Meaning of the Body: Aesthetics of Human Understanding*. University of Chicago Press.

Johnson, Mark L., and Don M. Tucker. 2021. *Out of the Cave: A Natural Philosophy of Mind and Knowing*. MIT Press.

Jones, Cameron R., and Benjamin K. Bergen. 2024. “People Cannot Distinguish GPT-4 from a Human in a Turing Test.” arXiv:2405.08007. Preprint, arXiv, May 9. https://doi.org/10.48550/arXiv.2405.08007.

Jones, Nicola. 2026. “How AI Is Shaping the War in Iran — and What’s next for Future Conflicts.” *Nature*, ahead of print, March 5. https://doi.org/10.1038/d41586-026-00710-w.

Keulartz, Jozef, Maartje Schermer, Michiel Korthals, and Tsjalling Swierstra. 2004. “Ethics in Technological Culture: A Programmatic Proposal for a Pragmatist Approach.” *Science, Technology, & Human Values* 29 (1): 3–29. https://doi.org/10.1177/0162243903259188.

Knight, Will. 2018. “The US Military Is Funding an Effort to Catch Deepfakes and Other AI Trickery.” *MIT Technology Review*, May 23. https://www.technologyreview.com/s/611146/the-us-military-is-funding-an-effort-to-catch-deepfakes-and-other-ai-trickery/.

Köbis, Nils C., Barbora Doležalová, and Ivan Soraperra. 2021. “Fooled Twice: People Cannot Detect Deepfakes but Think They Can.” *iScience* 24 (11): 103364. https://doi.org/10.1016/j.isci.2021.103364.

Koopman, Colin. 2009. *Pragmatism as Transition: Historicity and Hope in James, Dewey, and Rorty*. Columbia University Press.

——. 2011. “Rorty’s Linguistic Turn: Why (More Than) Language Matters to Philosophy.” *Contemporary Pragmatism* 8 (1): 61–84. https://doi.org/10.1163/18758185-90000183.

——. 2014. “Conduct Pragmatism: Pressing Beyond Experientialism and Lingualism.” *European Journal of Pragmatism and American Philosophy* VI (2): 2. https://doi.org/10.4000/ejpap.313.

Korsgaard, Christine M. 2018. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford University Press.

Korshunov, Pavel, and Sébastien Marcel. 2020. “Deepfake Detection: Humans vs. Machines.” *arXiv:2009.03155*. http://arxiv.org/abs/2009.03155.

Kripke, Saul A. 1980. *Naming and Necessity*. Harvard University Press.

Kroes, Peter. 2012. *Technical Artefacts: Creations of Mind and Matter: A Philosophy of Engineering Design*. Springer

Lake, Brenden M., and Gregory L. Murphy. 2023. “Word Meaning in Minds and Machines.” *Psychological Review* 130 (2): 401–31. https://doi.org/10.1037/rev0000297.

Lee, Jiyoung, and Soo Yun Shin. 2022. “Something That They Never Said: Multimodal Disinformation and Source Vividness in Understanding the Power of AI-Enabled Deepfake News.” *Media Psychology* 25 (4): 531–46. https://doi.org/10.1080/15213269.2021.2007489.

Lecheler, Sophie, Sanne Kruikemeier, and Yael de Haan. 2019. “The Use and Verification of Online Sources in the News Production Process.” In *Journalism and Truth in an Age of Social Media*, edited by James E. Katz and Kate K. Mays. Oxford University Press.

Leibowicz, Claire R., Sean McGregor, and Aviv Ovadya. 2021. “The Deepfake Detection Dilemma: A Multistakeholder Exploration of Adversarial Dynamics in Synthetic Media.” *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, July 21, 736–44. https://doi.org/10.1145/3461702.3462584.

Levine, Steven. 2012. “Brandom’s Pragmatism.” *Transactions of the Charles S. Peirce Society* 48 (2): 125–40. https://doi.org/10.2979/trancharpeirsoc.48.2.125.

——. 2019. *Pragmatism, Objectivity, and Experience*. Cambridge University Press.

Li, Yuezun, Ming-Ching Chang, and Siwei Lyu. 2018. "In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking." *arXiv,* June 11. http://arxiv.org/abs/1806.02877.

Liao, Yanlin, and Zihan Niu. 2025. "A Moderate Defense for Visual Arguments: Dynamic Existentialism and the Expansion of Argumentation Theory." *Informal Logic* 45 (1): 1–34. https://doi.org/10.22329/il.v45i1.8965.

Maddocks, Sophie. 2020. "'A Deepfake Porn Plot Intended to Silence Me': Exploring Continuities between Pornographic and 'Political' Deep Fakes." *Porn Studies* 7 (4): 415–23. https://doi.org/10.1080/23268743.2020.1757499.

Mallory, Fintan. 2023. "Fictionalism about Chatbots." *Ergo an Open Access Journal of Philosophy* 10. https://doi.org/10.3998/ergo.4668.

Manca, Danilo. 2022. "The Situated Mind and the Space of Reasons." *European Journal of Pragmatism and American Philosophy* XIV (2): 2. https://doi.org/10.4000/ejpap.3000.

Marchettoni, Leonardo. 2025. "There Are No Aliens in the Space of Reasons." *Philosophy & Technology* 38 (3): 1–13. https://doi.org/10.1007/s13347-025-00917-1.

Marin, Lavinia. 2022. "How to Do Things with Information Online. A Conceptual Framework for Evaluating Social Networking Platforms as Epistemic Environments." *Philosophy & Technology* 35 (3): 77. https://doi.org/10.1007/s13347-022-00569-5.

McDowell, John. 1994. *Mind and World*. Harvard University Press.

——. 2006. "Conceptual Capacities in Perception." In McDowell 2009.

——. 2007. "What Myth?" *Inquiry* 50 (4): 338–51. https://doi.org/10.1080/00201740701489211.

——. 2008. "Avoiding the Myth of the Given." In McDowell 2009.

——. 2009. *Having the World in View: Essays on Kant, Hegel, and Sellars*. Harvard University Press.

McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. MIT Press.

Millière, Raphaël. 2022. "Deep Learning and Synthetic Media." *Synthese* 200 (3): 1–27. https://doi.org/10.1007/s11229-022-03739-2.

Mirsky, Yisroel, and Wenke Lee. 2020. "The Creation and Detection of Deepfakes: A Survey." *arXiv,* May 12. http://arxiv.org/abs/2004.11138.

Misak, Cheryl. 2014. “Language and Experience for Pragmatism.” *European Journal of Pragmatism and American Philosophy* VI (2): 2. https://doi.org/10.4000/ejpap.295.

Montemayor, Carlos, Jodi Halpern, and Abrol Fairweather. 2022. “In Principle Obstacles for Empathic AI: Why We Can’t Replace Human Empathy in Healthcare.” *AI & SOCIETY* 37 (4): 1353–59. https://doi.org/10.1007/s00146-021-01230-z.

Neurath, Otto. (1921) 1973. “Anti-Spengler.” In *Empiricism and Sociology*, edited by Marie Neurath and Robert S. Cohen. Springer Netherlands.

Nguyen, C. Thi. 2020. “Echo Chambers and Epistemic Bubbles.” *Episteme* 17 (2): 141–61. https://doi.org/10.1017/epi.2018.32.

Nietzsche, Friedrich. (1873) 2005. “On Truth and Lies in a Nonmoral Sense.” In *Truth: Engagements Across Philosophical Traditions*, edited by José Medina and Wood, David. John Wiley & Sons

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

OpenAI. 2022. “Introducing ChatGPT.” November 30. https://openai.com/index/chatgpt/.

Paris, Britt, and Joan Donovan. 2019. “Deepfakes and Cheap Fakes.” *Data & Society*, September 18. https://datasociety.net/output/deepfakes-and-cheap-fakes/.

Pariser, Eli. 2011. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin.

Peele, Jordan, dir. 2018. *You Won’t Believe What Obama Says In This Video!* 😉. BuzzFeedVideo. https://www.youtube.com/watch?v=cQ54GDm1eL0.

Piantadosi, Steven T., and Felix Hill. 2022. “Meaning without Reference in Large Language Models.” *arXiv,* August 12. https://doi.org/10.48550/arXiv.2208.02957.

Powers, Thomas M., and Jean-Gabriel Ganascia. 2020. “The Ethics of the Ethics of AI.” In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale, and Sunit Das. Oxford University Press.

Price, Huw. 2011. *Naturalism Without Mirrors*. Oxford University Press.

Putnam, Hilary. 1975. “The Meaning of ‘Meaning’.” In *Mind, Language and Reality: Philosophical Papers, Volume 2*. Cambridge University Press

——. 1995. *Pragmatism: An Open Question*. Wiley-Blackwell.

——. 2004. “The Three Enlightenments.” In *Ethics without Ontology*. Harvard University Press.

Qian, Sijia, Cuihua Shen, and Jingwen Zhang. 2023. “Fighting Cheapfakes: Using a Digital Media Literacy Intervention to Motivate Reverse Search of out-of-Context Visual Misinformation.” *Journal of Computer-Mediated Communication* 28 (1): zmac024. https://doi.org/10.1093/jcmc/zmac024.

Quine. Willard van Orman. 1948. “On What There Is” In Quine 1961.

——. 1951. “Two Dogmas of Empiricism” In Quine 1961.

——. 1961. *From a Logical Point of View.* 2nd Revised Edition. With Richard M. Martin. Harvard University Press.

——. 1992. *The Pursuit of Truth*. Harvard University Press.

——. (1960) 2013. *Word and Object*. New ed. MIT Press.

Regan, Tom. 1983. *The Case for Animal Rights*. University of California Press.

Reuters. 2022. “Zelenskiy: 137 People Have Died in the Russian Invasion – Video.” World News. *The Guardian*, February 25. https://www.theguardian.com/world/video/2022/feb/25/zelenskiy-137-people-have-died-in-the-russian-invasion-video.

Rini, Regina. 2020. “Deepfakes and the Epistemic Backstop.” *Philosophers’ Imprint* 20 (24): 1–16.

Roque, Georges. 2015. “Should Visual Arguments Be Propositional in Order to Be Arguments?” *Argumentation* 29 (2): 177–95. https://doi.org/10.1007/s10503-014-9341-3.

Rorty, Richard. 1967. Introduction to *The Linguistic Turn: Recent Essays in Philosophical Method*, ed. Richard Rorty, 1-39. University of Chicago Press.

——. 1970. “Strawson’s Objectivity Argument” In Rorty 2014.

——. 1971. “Verification and Transcendental Arguments.” In Rorty 2014.

——. 1972. “The World Well Lost.” In Rorty 1982.

—— . 1977. “Dewey’s Metaphysics.” In Rorty 1982.

——. 1979. *Philosophy and the Mirror of Nature*. Princeton University Press.

——. 1980. “Pragmatism, Relativism, and Irrationalism.” In Rorty 1982.

——. 1982. *Consequences of Pragmatism*. University of Minnesota Press.

——. 1984. “Solidarity or Objectivity?” In Rorty 1991b.

——. 1985. “Texts and Lumps” In Rorty 1991b.

——. 1986. “Pragmatism, Davidson, and truth” In Rorty 1991b.

——. 1987a. “Non-reductive physicalism” In Rorty 1991b.

——. 1987b. “Unfamiliar Noises: Hesse and Davidson on Metaphor.” In Rorty 1991b.

——. 1989. *Contingency, Irony, and Solidarity*. Cambridge University Press.

——. 1991a. *Essays on Heidegger and Others*. Cambridge University Press.

——. 1991b. *Objectivity, Relativism, and Truth*. Cambridge University Press.

——. 1991c. “Inquiry as recontextualization” In Rorty 1991b.

——. 1996. “Something to Steer By.” Review of *John Dewey and the High Tide of American Liberalism*, by Alan Ryan. Philosophy & Law. *London Review of Books*, June 20. https://www.lrb.co.uk/the-paper/v18/n12/richard-rorty/something-to-steer-by.

——. 2000a. “Response to Allen.” In Brandom 2000b.

——. 2000b. “Response to Brandom.” In Brandom 2000b.

——. 2000c. “Response to Ramberg.” In Brandom 2000b.

——. 2004. “A pragmatist view of contemporary analytic philosophy” In Rorty 2007c.

——. 2007a. “Naturalism and quietism” In Rorty 2007c.

——. 2007b. “Wittgenstein and the linguistic turn.” In Rorty 2007c.

——. 2007c. *Philosophy as Cultural Politics*. Cambridge University Press.

——. 2014. *Mind, Language, and Metaphilosophy: Early Philosophical Papers*. Edited by Stephen Leach and James Tartaglia. Cambridge University Press.

——. 2021. *Pragmatism as Anti-Authoritarianism*. Edited by Eduardo Mendieta. Harvard University Press.

Rose, Steve. 2026. “The Slopaganda Era: 10 AI Images Posted by the White House – and What They Teach Us.” US News. *The Guardian*, January 29. https://www.theguardian.com/us-

news/2026/jan/29/the-slopaganda-era-10-ai-images-posted-by-the-white-house-and-what-they-teach-us.

Rössler, Andreas, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. 2019. "FaceForensics++: Learning to Detect Manipulated Facial Images." *arXiv*, August 26. http://arxiv.org/abs/1901.08971.

Russell, Bertrand. (1909) 2005. "William James's Conception of Truth." In *Truth: Engagements Across Philosophical Traditions*, edited by José Medina and David Wood. Wiley-Blackwell.

Sachs, Carl B. 2014. *Intentionality and the Myths of the Given: Between Pragmatism and Phenomenology*. Routledge.

Sambrotta, Mirco. 2025. "LLMs and the Logical Space of Reasons." *Minds and Machines* 35 (4): 46. https://doi.org/10.1007/s11023-025-09751-y.

Schoenherr, Jordan Richard. 2023. "The Disturbing Trend of State Media Use of Deepfakes." *The Conversation*, April 12. http://theconversation.com/the-disturbing-trend-of-state-media-use-of-deepfakes-203368.

Scruton, Roger. 1981. "Photography and Representation." *Critical Inquiry* 7 (3): 577–603.

Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–24. https://doi.org/10.1017/S0140525X00005756.

Sellars, Wilfrid. 1950. "Language, Rules, and Behavior." In *John Dewey: Philosopher of Science and Freedom*, edited by Sidney Hook. Dial Press.

———. 1956. "Empiricism and the Philosophy of Mind" In Sellars 1997.

———. 1962. "Philosophy and the Scientific Image of Man" In Sellars 1963.

———. 1963. *Science, Perception and Reality*. Ridgeview Publishing Company.

———. 1965. "Scientific Realism or Irenic Instrumentalism: A Critique of Nagel and Feyerabend on Theoretical Explanation." In *Boston Studies in the Philosophy of Science, Vol. II*, edited by Robert Cohen and Max Wartofsky. Humanities Press.

———. 1997. *Empiricism and the Philosophy of Mind*. Edited by Robert Brandom. With Richard Rorty. Harvard University Press.

Shanahan, Murray, Kyle McDonell, and Laria Reynolds. 2023. "Role Play with Large Language Models." *Nature* 623 (7987): 493–98. https://doi.org/10.1038/s41586-023-06647-8.

Sharma, Vishal Kumar, Rakesh Garg, and Quentin Caudron. 2024. "A Systematic Literature Review on Deepfake Detection Techniques." *Multimedia Tools and Applications* 84 (20): 22187–229. https://doi.org/10.1007/s11042-024-19906-1.

Simonelli, Ryan. 2022. "Sellars's Ontological Nominalism." *European Journal of Philosophy* 30 (3): 1041–61. https://doi.org/10.1111/ejop.12729.

Sleeper, R. W. 1985. "Rorty's Pragmatism: Afloat in Neurath's Boat, but Why Adrift?" *Transactions of the Charles S. Peirce Society* 21 (1): 9–20.

Somoray, Klaire, and Dan J. Miller. 2023. "Providing Detection Strategies to Improve Human Detection of Deepfakes: An Experimental Study." *Computers in Human Behavior* 149 (C). https://doi.org/10.1016/j.chb.2023.107917.

Søgaard, Anders. 2023. "Grounding the Vector Space of an Octopus: Word Meaning from Raw Text." *Minds and Machines* 33 (1): 33–54. https://doi.org/10.1007/s11023-023-09622-4.

Strawson, Peter. F. 1966. *Bounds of Sense*. Routledge.

Sundar, S Shyam. 2008. "The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility." In *Digital Media, Youth, and Credibility*, edited by Miriam J. Metzger and Andrew J. Flanagan. Eries on Digital Media and Learning. MIT Press.

Symons, John, and Ramón Alvarado. 2022. "Epistemic Injustice and Data Science Technologies." *Synthese* 200 (2): 1–26. https://doi.org/10.1007/s11229-022-03631-z.

Taeb, Maryam, and Hongmei Chi. 2022. "Comparison of Deepfake Detection Techniques through Deep Learning." *Journal of Cybersecurity and Privacy* 2 (1): 89–106. https://doi.org/10.3390/jcp2010007.

Tanno, Sophie. 2022. "Ukraine-Russia Lunchtime Briefing: Five Developments as UK Sends Sky Sabre Missiles to Poland." *The Telegraph*, March 17. https://www.telegraph.co.uk/world-news/2022/03/17/ukraine-russia-lunchtime-briefing-five-developments-uk-sends/.

Thomasson, Amie. 2007. *Ordinary Objects*. Oxford University Press.

Thompson, Stuart A., and Tiffany Hsu. 2023. "How Easy Is It to Fool A.I.-Detection Tools?" *The New York Times*, June 28. https://www.nytimes.com/interactive/2023/06/28/technology/ai-detection-midjourney-stable-diffusion-dalle.html.

Tomasello, Michael. 1999. *The Cultural Origins of Human Cognition*. Harvard University Press.

Tripodi, Francesca B., Lauren C. Garcia, and Alice E. Marwick. 2023. "'Do Your Own Research': Affordance Activation and Disinformation Spread." *Information, Communication & Society* August 30: 1–17. https://doi.org/10.1080/1369118X.2023.2245869.

Vaccari, Cristian, and Andrew Chadwick. 2020. "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News." *Social Media + Society* 6 (1): 2056305120903408. https://doi.org/10.1177/2056305120903408.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press.

——. 2024. *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*. Oxford University Press.

Van Den Eede, Yoni. 2010. "'Conversation of Mankind' or 'Idle Talk'?: A Pragmatist Approach to Social Networking Sites." *Ethics and Information Technology* 12 (2): 195–206. https://doi.org/10.1007/s10676-010-9223-9.

Varughese, Jobit. 2025. "What Is a Multimodal LLM (MLLM)?. *IBM*, July 30. https://www.ibm.com/think/topics/multimodal-llm.

Verbeek, Peter-Paul. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. University of Chicago Press.

Voparil, Chris. 2021. *Reconstructing Pragmatism: Richard Rorty and the Classical Pragmatists*. Oxford University Press.

Wakefield, Jane. 2024. "Tackling Deepfakes 'Has Turned into an Arms Race.'" *BBC News*, March 27. https://www.bbc.com/news/business-68549609.

Walton, Kendall L. 1984. "Transparent Pictures: On the Nature of Photographic Realism." *Critical Inquiry* 11 (2): 246–77.

Waltzer, Tal, Celeste Pilegard, and Gail D. Heyman. 2024. "Can You Spot the Bot? Identifying AI-Generated Writing in College Essays." *International Journal for Educational Integrity* 20 (1): 11. https://doi.org/10.1007/s40979-024-00158-3.

Watson, David S., Jakob Mökander, and Luciano Floridi. 2025. "Competing Narratives in AI Ethics: A Defense of Sociotechnical Pragmatism." *AI & SOCIETY* 40 (5): 3163–85. https://doi.org/10.1007/s00146-024-02128-2.

Weikmann, Teresa, and Sophie Lecheler. 2023. "Visual Disinformation in a Digital Age: A Literature Synthesis and Research Agenda." *New Media & Society* 25 (12): 3696–713. https://doi.org/10.1177/14614448221141648.

——. 2024. “Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network.” *Digital Journalism* 12 (10): 1505–22. https://doi.org/10.1080/21670811.2023.2194665.

Weizenbaum, Joseph. 1966. “ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine.” *Communications of the ACM* 9 (1): 36–45. https://doi.org/10.1145/365153.365168.

West, Jevin, and Carl Bergstrom. 2019. “Which Face Is Real?” Which Face Is Real. Accessed April 28, 2020. http://www.whichfaceisreal.com/index.php.

Williams, Iwan, and Tim Bayne. 2024. “Chatting with Bots: AI, Speech Acts, and the Edge of Assertion.” *Inquiry*, December 10, 1–24. https://doi.org/10.1080/0020174X.2024.2434874.

Wineburg, Sam, and Sarah McGrew. 2019. “Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital Information.” *Teachers College Record* 121 (11): 1–40.

Winner, Langdon. 2020. *The Whale and the Reactor: A Search for Limits in an Age of High Technology, Second Edition*. 2nd ed. University of Chicago Press.

Wittgenstein, Ludwig. (1953) 2009. *Philosophical Investigations*. Rev. 4th ed. Translated by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Wiley-Blackwell.

——. 1958. *The Blue and Brown Books*. Blackwell.

Wolfram, Stephen. 2023. “What Is ChatGPT Doing … and Why Does It Work?” *Stephen Wolfram Writings*, February 14. https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.

Yang, Xin, Yuezun Li, and Siwei Lyu. 2018. “Exposing Deep Fakes Using Inconsistent Head Poses.” *arXiv [Cs]*, November 13. http://arxiv.org/abs/1811.00661.

Zelenskyy, Volodymyr. 2022. “Те, що окупанти роблять Харкову, Охтирці, Києву, Одесі та іншим містам і селищам, тягне на міжнародний трибунал.” Facebook, February 26. https://www.facebook.com/share/v/1JbCrNsYzW/.