The Structure of Iconic Representation

ANDREW Y. LEE | University of Toronto, Philosophy
JOSHUA MYERS | Western University, Philosophy
GABRIEL OAK RABIN | New York University Abu Dhabi, Philosophy

Abstract

This paper develops a theory of iconic representation. We begin by examining prior proposals that appeal to the parts principle, according to which iconic representations are those where parts of the vehicle represent parts of the content. Then we develop our positive theory, according to which iconic representations are *analog locative structures*—collections of analog representations organized into what we call a 'locative structure'. Along the way, we clarify the relationship between the iconic and the analog, precisify the contested notion of a "functional space" from cognitive science, explain how our theory can be applied to mental representations, and generalize our theory beyond paradigms of iconic representation.

Introduction

"A picture is worth a thousand words." This adage is familiar to nearly everyone, and to some it may feel like an obvious truism. But behind the adage lies a philosophically interesting observation: a picture and a collection of words can be used to represent the very same thing, but they do so in very different ways. Using more theoretical language, the picture represents **iconically**, while the collection of words does not.

There's no agreement on what exactly it takes for a representation to be iconic.[1] But there tends to be agreement that the term 'iconic' denotes—

[1] There are two distinct but related threads of research labeled 'iconic representation.' The first, which traces its roots back to the imagery debate in cognitive science, takes iconic representations to be broadly imagistic or pictorial in nature (Fodor [2008], Quilty-Dunn [2016], Clarke [2022]). The second, which largely stems from linguistics, takes iconic representations to be those that bear a natural or non-arbitrary correspondence to what they represent (Schlenker, Lamberton, & Santoro [2013], Kuhn & Aristodemo [2017], Greenberg [2023]). The first kind is narrower than the second: a mercury thermometer isn't pictorial in nature, but still involves non-arbitrary correspondence between vehicles and contents. We use 'iconic' to denote the narrower kind and 'analog' to denote the broader

or should denote—a more general representational kind than merely pictures. Alongside pictures (and nearby kinds, such as photographs, videos, and drawings), the category of the iconic may reasonably be taken to also include sculptures, dioramas, audio recordings, mental images, and visual experiences.[2]

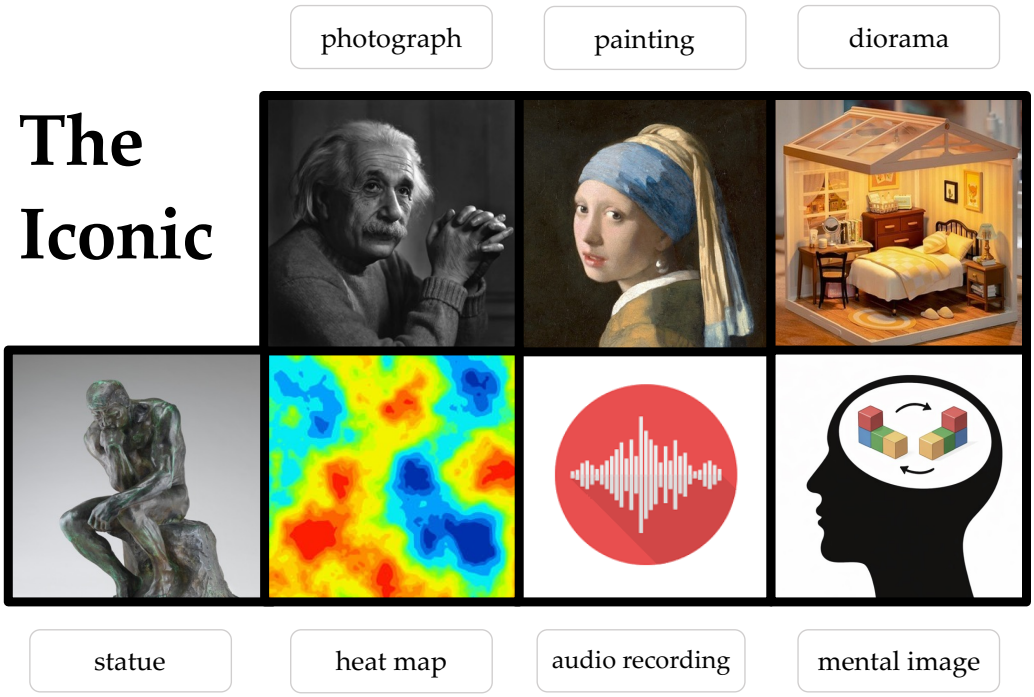

FIGURE 1: A sample of iconic representations.

These contrast with representations consisting of words or symbols, which we will call **discursive**, as well as representations such as graphs,

kind. Those who prefer a different use of 'iconic' should feel free to substitute 'imagistic' or 'pictorial' instead.

[2] Relatedly, there is a significant literature on the nature of *depiction*, which is typically characterized as the type of representing that pictures do (Abell [2009], Hopkins [1995], Hyman [2012], Peirce [1982], Schier [1986], Wollheim [1998]). This literature focuses mostly on visual artworks, rather than the wider class of representational systems that we are targeting. Furthermore, this literature mostly focuses on individual representations, rather than representational systems. For these reasons, we will leave open the relationship between depiction and iconic representation.

diagrams, charts, and maps, which often have both iconic and discursive elements.

The goal of this paper is to identify what exactly it takes for a representation to be iconic. The basic idea behind our proposal is that iconic representation involves three ingredients. They are:

MOLECULARITY: Iconic representations have multiple parts.
ANALOGICITY: Iconic representations represent analogically.
LOCATIVITY: Iconic representations have locative structure.

We will spend a good deal of time discussing each of these principles, including a detailed explanation of what we mean by "locative structure." Our core proposal, expressed in a slogan, is that iconic representations are **analog locative structures**.

A preliminary point: it's important to distinguish between individual representations and representational systems. A photograph of Jerry Fodor is an individual representation, but it is a member of the representational system we will call *Color Photographs*. A representational system is a set of possible vehicles, a set of possible contents, and an interpretation function that maps vehicles to contents. For example, a photograph is itself a vehicle, the scene it represents is a content, and its interpretation function maps parts and features of the photograph to parts and features of the scene. On our view, iconicity is primarily a property of representational systems; individual representations count as iconic in virtue of belonging to an iconic system. Nevertheless, for simplicity of exposition, we will often use individual representations to illustrate the nature of the representational systems to which they belong.

Here's the plan. **§1** examines various iterations of the parts principle, the idea that iconic representations are those where parts of the vehicle represent parts of the content. **§2** develops the notion of "locative structure," a central component of our view. **§3** develops our positive theory, where the core idea is that iconic representations are locatively structured analog representations. **§4** explains how to generalize our theory to more complex representational systems. **§5** explains how to apply our

theory to iconic mental representations. **§6** applies the concepts in our theory to illustrate representational joints beyond the iconic.

§1 The Parts Principle

To start, we will address various iterations of the parts principle. The general idea behind the parts principle is that iconic representations are those where parts of the vehicle represent parts of the content. We focus on the parts principle for two reasons. First, it has been one of the most influential ideas in the literature on iconic representation.[3] Second, while we will argue that even the best iterations of the parts principle are inadequate as general theories of iconic representation, we believe that the parts principle still tracks an important truth. In what follows, we examine various iterations of the parts principle and draw lessons from their counterexamples to develop our positive theory of iconic representation.

The parts principle was originally proposed in the philosophical literature by Elliot Sober [1976], but it finds its canonical statement in the work of Jerry Fodor [2008: 173]: "If P is a picture of X, then parts of P are pictures of parts of X."[4] Here's a natural way of generalizing this idea in order to yield an analysis of what it is for a representation to be iconic (let *v* be a vehicle and *c* be a content):

Simple Parts Principle
v is an iconic representation of *c* iff every part of *v* is a representation of some part of *c*.[5]

[3] Proponents of the parts principle include Carey [2009], Kulvicki [2015], Quilty-Dunn [2016, 2020a, 2020b], Green & Quilty-Dunn [2021], and Clarke [2020]. For critical discussions, see Balog [2009], Burge [2018], Clarke [2022].

[4] The parts principle can be taken as necessary, sufficient, or necessary and sufficient for iconic representation. Our goal is to offer an analysis of iconic representation, not merely to propose a one way conditional. Because of this, we will interpret the parts principle and all the variants we discuss as necessary and sufficient conditions for iconic representation.

[5] A variant on the Simple Parts Principle says that iconic representations are those where every part *iconically* represents a part of the content. This variant better mirrors Fodor's formulation of the picture principle, which requires that the parts not only represent, but themselves be pictures. However, such a variant is in danger of circularity, since it uses the

FIGURE 2: A photograph of Jerry Fodor, alongside its left half.

The idea is intuitive enough. Consider a photograph of Jerry Fodor. If you cut the photograph in half, the left half of the representation represents the left half of Fodor. Further cuts result in successively smaller photographs that represent successively smaller parts of Fodor. By contrast, cutting the word 'Fodor' in half doesn't result in representations of parts of Fodor—the 'Fod' in 'Fodor' doesn't represent anything. The Simple Parts Principle thus seems to offer a promising way of distinguishing between the iconic and the non-iconic.

An immediate problem is that the Simple Parts Principle is trivially satisfied by representations that lack parts altogether. Call these **atomic** representations, in contrast to **molecular** representations that have parts. Atomic representations will always satisfy the Simple Parts Principle since their only vehicular part (the improper part corresponding to the whole vehicle) represents their only content part (the improper part corresponding to the whole content). But counting all atomic representations as iconic would be far too permissive. Suppose we were to use the '●' symbol to represent Fodor. The fact that this representation lacks parts should not automatically make it iconic. Moving forward, we will

term 'iconic' to define 'iconic'. Furthermore, our objection to the Simple Parts Principle will apply to this variant principle as well.

assume that iconic representations must be molecular. Presumably this is what proponents of the parts principle had in mind all along.

Even so, there remains another problem for the Simple Parts Principle: not all parts of iconic representations are themselves representations. Each individual atom that makes up the photograph of Fodor is a part of the photograph, but the individual atoms are not representations of parts of Fodor.[6] If there are iconic representations in the brain, then they have neurons as parts, but individual neurons—let alone their subcellular constituents—are unlikely to be representations at all. In general, any representation will eventually be composed of some parts that are not themselves representations.[7] To circumvent this problem, one might appeal to the following amendment:

> **Syntactic Parts Principle**
> *v* is an iconic representation of *c* iff every syntactic part of *v* is a representation of some part of *c*.

This principle amends the previous one by specifying the relevant type of part: only *syntactic parts* matter. While atoms and neurons may be parts of iconic representations, they are (usually) not syntactic parts. The Syntactic Parts Principle thus avoids the counterexamples raised in the preceding paragraph. Nevertheless, the Syntactic Parts Principle is false.

One basic problem is that the Syntactic Parts Principle no longer captures the intended contrast between iconic and discursive representations. A common motivation for appealing to parts principles is the following sort of observation: while the left half of the photograph of

[6] Fodor [2008] mentions this problem in a footnote (*fn* 6) only to set it aside. He does not suggest how to revise the parts principle in response to this counterexample.

[7] The only possible exception would be if there are representations that have a gunky representational structure, where all parts have further parts (that themselves represent). But a theory of iconic representation should not be beholden to these sorts of controversial metaphysical commitments. Moreover, even if such gunky representations are possible, the Simple Parts Principle would still misclassify many paradigms of iconic representation, such as ordinary photographs, since such representations have atomic representational parts.

Fodor represents a part of Fodor, the left half of the word 'Fodor' doesn't represent anything. But suppose now that we restrict our focus to syntactic parts. The basic syntactic parts of sentences are words or morphemes, rather than individual letters: 'Fod' simply isn't a syntactic part of 'Fodor'. But this means that it's no longer clear that sentences of natural language fail to satisfy the Syntactic Parts Principle. If we consider the sentence 'Fodor wrote many philosophy papers', it seems that 'Fodor' and 'philosophy papers' both represent parts of what the sentence represents. Since sentences of natural language are paradigms of non-iconic representations, the Syntactic Parts Principle is inadequate.

One might object that while both iconic and discursive representations have syntactic parts that play a role in determining the content of the whole, iconic representations are distinguished by the fact that their syntactic parts each represent a part of the content (because discursive contents either lack parts altogether or have parts that do not correspond appropriately). Suppose, for example, that contents are sets of possible worlds. Then the content of the sentence 'Fodor wrote many philosophy papers' is a set of worlds. Since the contents of 'wrote' and 'many' won't be any part of that set, this move enables the proponent of the Syntactic Parts Principle to count sentences as non-iconic. However, this move requires appealing to a specific proposal according to which the contents of sentences are sets of possible worlds. Many philosophers of language instead favor a structured propositions view, according to which propositions have structures that correspond to the syntactic structure of the natural language sentences that express them.[8] If a structured propositions view is correct, then the Syntactic Parts Principle no longer excludes sentences from counting as iconic. This is problematic: the correct theory of iconic representation should not hinge on whether the structured propositions theory of linguistic content turns out to be correct.

Setting sentences aside, there are other counterexamples to the Syntactic Parts Principle. Consider a list of names that represents the members of an elementary school class. Each syntactic part of the list (each name) represents a part of the class (a student), so the Syntactic Parts

[8] See Salmon [1986], Soames [1987], and King [2007].

Principle is satisfied. Nevertheless, lists of names are not iconic. Or consider a bank of digital clocks that represents the current times in multiple cities (Tokyo, Paris, New York, etc.). Each syntactic part of the clock (each sequence of four numerals, such as '16:04') represents a part of the content (the time in a city). Nevertheless, this bank of clocks is not iconic.

To address these kinds of counterexamples, one might appeal to *relations between parts* to demarcate the iconic from the non-iconic. Compare a class photograph (which is iconic) to a list of names (which is not). In the photograph, relations between parts are mirrored by relations between contents. If the part that represents Alice is to the right of the part that represents Benito, then Alice is represented as to the right of Benito. By contrast, in the list, spatial relations between names have no representational significance: the fact that 'Alice' occurs above 'Benito' has no implications for the content of the list. In general, iconic representations seem to involve correspondence of structural relations in a way that discursive representations do not. These observations motivate the following principle:[9]

Syntactic Relations Principle
v is an iconic representation of *c* iff (**1**) every syntactic part of *v* is a representation of some part of *c*, and (**2**) every syntactic relation over the parts of *v* represents a semantic relation over the parts of *c*.

The Syntactic Relations Principle is an improvement: it prevents lists of names from counting as iconic. However, it is still subject to counterexamples. Consider *Seating Chart*, a representational system for the seat locations for the students in an elementary school class. The seating chart consists of twelve names, arranged into three rows of four. Each name represents the seat location for that student, and the arrangement of names on the seating chart corresponds to the arrangement of the students' seats:

[9] Similar principles are discussed by Kosslyn [1981], Tye [1991:44], Quilty-Dunn [2020] and Clarke [2022].

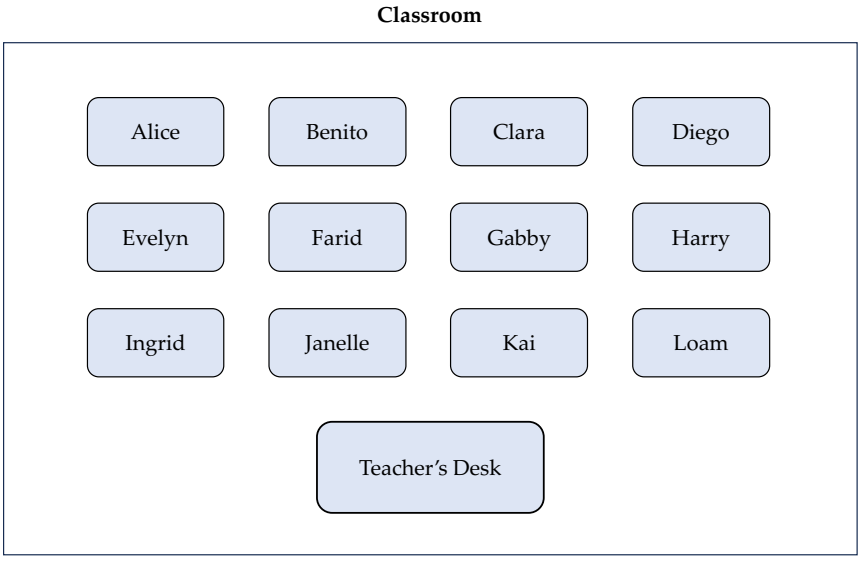

FIGURE 3: A seating chart.

Since each part of the seating chart (each name) represents a member of the class, and since spatial relations between names correspond to spatial relations between members of the class, *Seating Chart* satisfies the Syntactic Relations Principle. Yet *Seating Chart* is still missing an ingredient of iconicity; names written in natural language are paradigms of non-iconic representation. The general problem is that one can always take some arbitrary symbols (like names) and slap some relations on them so that the principle is satisfied. But iconic representations are not composed from arbitrary symbols, no matter how many syntactic relations those symbols stand in to each other.

At this point, it's natural to invoke the distinction between analog versus symbolic representation. Consider the following principle:

Analog Parts Principle
v is an iconic representation of *c* iff every syntactic part of *v* is an analog representation of some part of *c*.[10]

[10] We take analog representation to be first and foremost a matter of features, rather than parts. But we will say that a part represents analogically whenever all of its features represent analogically.

A number of theorists have thought that analog (as opposed to symbolic) representation is an important ingredient for iconicity.[11] To characterize the distinction, it's useful to start with some paradigm examples. Mercury thermometers, hand clocks, and pictures represent analogically. Digital thermometers, digital clocks, and words represent symbolically. Even when two representations represent the exact same things (such as mercury thermometers and digital thermometers), one can do so analogically and the other symbolically.

The exact nature of the difference between analog representation and symbolic representation is a matter of controversy. But a common thread that unites many theories is that analog representation is a matter of structure in the vehicles corresponding to structure in the contents.[12] Consider, for example, the difference between a mercury thermometer and a digital thermometer. In the former, greater mercury heights represent warmer temperatures and thus structure is mirrored between vehicles and contents. In the latter, there is no structural correspondence between digits and temperatures.

As another example, consider the difference between a hand clock and a digital clock. In hand clocks, the angle of the minute hand represents how much time has passed since the hour changed. Greater angles represent greater spans of time; again, structure is mirrored between vehicles and contents. By contrast, for digital clocks, there is no structural correspondence between the shapes of the numerals and the times that those numerals represent. For the purposes of this paper, we will simply take for granted the common idea that analog representation is a matter of the sort of structural correspondence described above.

[11] See Quilty-Dunn [2017: ch. 4], Clarke [2022], Block [2023: 222], Maley [2023: 130], and Langland-Hassan [2025].

[12] Different theories have different ways of cashing out the idea that analog representation involves structural correspondence. For example, Maley [2011], Beck [2019], Peacocke [2019], and Block [2023: 222] appeal to magnitudes mirroring magnitudes. Others, such as Kulvicki [2015], appeal to abstractions corresponding to abstractions. Our preferred theory is Lee, Myers, & Rabin [2023], who hold that analog representations use rules that map syntactic structural features onto semantic structural features. For the purposes of this paper, though, we remain agnostic between these different options.

The Analog Parts Principle, unlike the Syntactic Relations Principle, correctly excludes *Seating Chart* from counting as iconic. *Seating Chart* has two representational features: name and spatial position. Spatial position represents analogically, since spatial position on the seating chart corresponds to spatial position in the classroom. But names do not represent analogically: for example, there is no structural correspondence between the name 'Abdul' and the student that it represents. The system thus doesn't satisfy the Analog Parts Principle.

Although the Analog Parts Principle is an improvement on the Syntactic Relations Principle, it is still subject to counterexamples. Suppose you want to figure out the average temperature in your neighborhood during the summer. Every day at noon, you take a temperature reading using a single-use analog thermometer and toss it in a box. By the end of the summer, you have a box of thermometers representing an unordered set of temperatures that you can use to calculate the average temperature. Call this system *Box of Thermometers*:

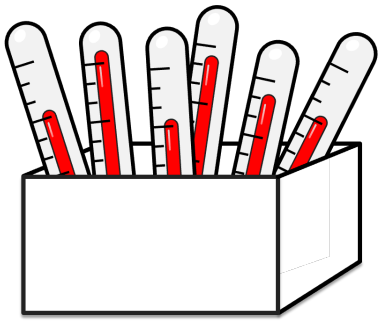

FIGURE 4: A box of thermometers.

This system satisfies the Analog Parts Principle, since each syntactic part (each individual thermometer) represents analogically. Nevertheless, *Box of Thermometers* is not iconic.

One might seek to combine the requirements in the Analog Parts Principle and the Syntactic Relations into a combination view that takes iconicity to require both analog parts and mirroring of syntactic relations. However, this natural line of thought is also defeated by *Box of Thermometers*. We just discussed how *Box of Thermometers* has analog parts:

the thermometers. But notice also that "has greater height of mercury than" relations between individual thermometers mirror "warmer temperature than" relations between the contents of those thermometers. Despite having both analog parts and mirrored syntactic relations, *Box of Thermometers* is not iconic. This means that even the combination view is inadequate.

We believe that the ideas motivating these various iterations of the parts principle are on the right track. The Simple Parts Principle is correct that iconic representations must have parts, and those parts must correspond to the parts of what is represented. The Syntactic Parts Principle is correct to focus on syntactic parts rather than all physical parts. The Syntactic Relations Principle is correct that in iconic representations, relations between parts must correspond to relations between what is represented. The Analog Parts Principle is correct that the parts of iconic representations must represent analogically. However, none of these principles quite identifies the target representational kind, and all of these principles are subject to counterexamples.

The aforementioned principles point towards the idea that iconic representations are, in some sense, collections of analog representations. But to be iconic, those collections of analog representations must also be structured in the right way. Compare *Box of Thermometers* to *Color Photographs*. Both representations involve collections of parts that represent analogically. But while *Box of Thermometers* represents a mere aggregate of temperatures, *Color Photographs* does more than represent a mere aggregate of colors—it also uses spatial relations to locate those colors in space. Only *Color Photographs* exhibits the kind of structure that is characteristic of iconic representation. The missing ingredient—we will soon argue—is what we call *locative structure.*

§2 Locative Structure

Our view is that iconic representations are locatively structured analog representations. Since locative structure is the most novel component of our

theory, we will spend a good deal of space explaining what locative structure is and why it is necessary for iconic representation.[13]

The idea of locative structure begins with the observation of an asymmetry. Consider *Color Photographs*. A photograph has many parts: its pixels. The representational state of the photograph can be captured by specifying the color and spatial position of each pixel. At first, it may be tempting to think of color and space as playing analogous representational roles: each represents a certain property of the content. But there's an important structural difference between the role played by color and the role played by space. The difference is illustrated by the following observation: any two pixels can share the same color, but no two pixels can share the same spatial position.

The asymmetry may also be illustrated by contrasting *Color Photographs* with *Box of Thermometers*. In a photograph, every pixel must occupy a unique spatial position. By contrast, no feature of *Box of Thermometers* works this way: every thermometer in the box can have the same height of mercury. If a system has a feature (or collection of features) that works this way—where each part of each representation must be unique with respect to that feature—then that system generates a *locative structure*. The idea of locative structure generalizes the representational role that spatial positions play in paradigm instances of iconic representation, such as photographs. In what follows, we develop these basic ideas in more systematic terms.

To simplify our initial presentation of the theory, we invoke three simplifying assumptions. The first—ATOMICITY—is that representations are wholly composed from atomic parts. The second—FEATURE COMPLETENESS—is that every part has a value along every feature. The third—COMPOSITIONALITY—is that the features of molecular representations are determined by the features of their parts.[14] These assumptions will

[13] Our notion of locative structure is partially inspired by Clark [2000: Ch. 4].

[14] ATOMICITY excludes representations that aren't ultimately composed of atomic parts (such as gunky representations). FEATURE COMPLETENESS excludes representations with parts that take values along some but not all features (such as "photographs" in which some pixels lack color altogether). COMPOSITIONALITY excludes representations that have emergent features—features of molecules that aren't reducible to features of their parts.

enable us to focus on explaining the core ideas behind our theory, rather than on covering the full space of possibilities. Because of this, these assumptions won't play a prominent role in the ensuing presentation. But we will return to them in §4, when we explain how to generalize our theory to systems that violate each of these assumptions.

2.1 Defining Locative Structure

To be iconic, a representation must have multiple parts. Those parts have **features**: the representationally relevant properties of the parts. Each part takes a specific **value** of a given feature: a specific way that a feature can be instantiated. The parts of a photograph are its pixels. The features are spatial position and color. The values are specific colors (scarlet, chartreuse) and specific spatial positions (the upper-right corner).[15] The total state of a photograph is determined by the values each part (pixel) takes with respect to the system's features.[16]

Every feature can be classified as either locative or qualitative, depending on whether every part must instantiate a unique collection of feature values. The basic idea behind the distinction is that two parts of a representation can share all their qualitative feature values, but cannot share all their locative feature values. Consider *Color Photographs*. If we think of this system as having two features—spatial position and color—then spatial position is locative and color is qualitative. Every pixel has a

[15] It's tempting to think of features as determinables and values as determinates. But the story is a bit more complicated. In one representational system, color could be a feature and red, yellow, and green the three available values. In another system, redness could be a feature and scarlet, burgundy, and raspberry the available values. Whether a property is a feature or a value depends on the syntax of the representational system in question, rather than the metaphysical structure of the property. Relatedly, token representations can instantiate different determinates while having identical values with respect to some feature; if the traffic light on Houston and Broadway shines scarlet while the traffic light one block down shines burgundy, both traffic lights have the same value: red.

[16] Strictly speaking, locative structure is best thought of as a property of representational systems (such as *Color Photographs*) rather than individual representations (such as the photograph of Fodor). But for simplicity of exposition, we will freely ascribe locative structures to individual representations over the rest of the paper.

unique spatial position, but there is no requirement that pixels have unique colors.

More precisely, a feature is **locative** just in case it's a member of a **locative set**, which we define as a minimal set of features such that every part takes a unique combination of values on those features:[17]

> **Definition: 'Locative Feature'**
> a feature that's a member of a *locative set*: a minimal set of features such that within every representation of the system, each part takes a unique combination of values on those features.

Equivalently, no two parts can share all their locative feature values. Any feature that isn't locative is **qualitative**.

In simpler contexts, it will be safe for the reader to think of locative features as those that aren't sharable. When a representational system has a locative set that contains only one feature, its locative status consists solely in the fact that every part has a unique value on that feature. *Color Photographs* plausibly operates this way. It has a one-member locative set containing the spatial position feature. No two pixels can share spatial position. Many pixels can share color.

Appealing to locative sets allows us to handle more complicated systems with multiple locative features. Suppose we instead construed *Color Photographs* as involving five features: hue, saturation, brightness, x-coordinate, and y-coordinate. On this construal, two pixels can share x-coordinate. But x-coordinate is locative because it forms a locative set with y-coordinate. The set is locative because every part has a unique combination of x- and y-coordinate values.

[17] There are two senses of 'minimal' at work in this definition. First, the *set* must be minimal, in the sense that no proper subset of a locative set can be a locative set. Second, the *feature* must be minimal, in that either all its sub-features are locative or all its sub-features are qualitative. The basic purpose of these clauses is to prevent "qualitative free-riders" (for example, to prevent color from counting as locative in *Color Photographs* because no two pixels can share both spatial position *and* color).

This illustrates two important points. First, the mere fact that a feature value can be shared between parts doesn't entail that it is qualitative. In order to determine whether a system has locative structure, one cannot merely look at each feature on its own. One instead needs to look at how features combine. Second, our definition of 'locative feature' works regardless of how finely we individuate features. In *Color Photographs,* spatial position comes out as locative regardless of whether we think of it as a single feature or as comprising distinct *x*- and *y*-coordinate features. More generally, if a feature is locative, then breaking it into multiple sub-features will still yield locative features.

Our definitions also work whether we think of features as monadic or relational (or as a combination of both). For example, imagine a photographic system where the relations left-of, right-of, above, and below are the fundamental spatial features. These relational features will still form a locative set, since no two pixels can bear all the same spatial relations to other pixels.[18]

As already discussed, our preferred notion of part is that of a syntactic part. Thus, we take the relevant parts to be determined by the syntax of the representational system under consideration. Given this, when we say that distinct parts of a representation *cannot* be identical with respect to all their locative feature values, we mean that the system's syntactic constraints (rather than physical or metaphysical constraints) prohibit such possibilities.

Consider, for example, how two pixels cannot share the same spatial position in a photograph. In principle, one could cut out the upper-right pixel and place it on top of the upper-left pixel, resulting in two pixels sharing the upper-left-most location. Then two pixels would, in a sense, share the same spatial position (at least with respect to their *x*- and *y*-coordinates). But it's obvious that the resulting entity isn't a well-formed representation of *Color Photographs*, in the same way that '& → PQ' isn't a well-formed formula of propositional logic. This example illustrates how

[18] We defined a feature as *locative* exactly when it's a member of a *locative set*, which is a minimal set of features such that distinct parts of a representation cannot share all the values they take of those features. This definition applies equally well to relational features.

it's the syntax of the system that makes the resulting representation ill-formed, not the laws of physics.

For a representational system to have **locative structure** is for it to have at least one locative feature. Representations with locative structure have a special quality: every part of the representation occupies a unique position in that structure. The uniqueness of these positions is generated by the fact that no two parts can take the same values with respect to all their locative features.[19] The locative feature values taken by one part constrain the possibilities for the feature values of the other parts. For example, once one pixel is located at $x = 0$, $y = 0$ (the lower-left corner of the photograph), no other pixel can take that combination of x- and y-coordinates. Qualitative features generate no such constraint, and qualitative features don't generate an abstract structure in which parts occupy unique positions.[20]

2.2 Locative Structure and Physical Space

The term 'locative' may evoke connotations of physical space. But it's worth emphasizing that spatial features need not always be locative, and locative features need not always be spatial. If there were a system where each part had to have a unique color but where two parts could share the same spatial position, color would be locative and spatial position qualitative.

---

[19] If a representational system doesn't permit representations with multiple parts, then trivially, no two parts take the same values with respect to all their features, and every part occupies a unique position in the feature space. As a result, such systems trivially satisfy our definition of locative structure (for an example, see *Mercury Thermometer* in §3.1). One could instead require that the system's representations have multiple parts in order to properly qualify as having locative structure, or one could appeal to how the system would treat additional parts were they to be added. We have no substantive objection to these proposals, though we opt for the simpler implementation without these additional qualifications.

[20] On our view, some iconic representations have only locative features, while others have both locative and qualitative features. The former might be thought of as analogous to blank canvases, and the latter as painted canvases. Some readers might choose to reserve the label 'iconic' for representations that have both types of features. We have no substantive objection to such a proposal so long as it also recognizes the existence of systems in which all features are locative. For that latter category they may use whatever label they feel appropriate.

Consider *Pinned Map*, a system of unlabeled maps marking the borders between the provinces and territories of Canada, with colored pins tacked inside the borders of states:

FIGURE 5: A pinned map.

Each color represents a distinct person, and a pin placed in a region means that the person represented by that pin's color lives in that province. In this example, color is locative because each pin must have a unique color. Conversely, spatial position is qualitative because two distinct pins can share their spatial position. For example, this will happen if both Dave and Jane live in Ontario.

While many paradigmatic iconic representations, such as *Color Photographs*, use physical space to play a locative role, our theory grounds iconicity in that role (locative structure) rather than its common realizer (physical space). This contrasts with approaches that build domain restrictions into the definition of iconic representation. For example, some

have proposed that iconic representations must represent physical space, or that they must represent visual or sensory features, or that they must represent concrete, non-abstract features.[21] However, we think that such domain restrictions are unmotivated.[22]

Consider *Climate Timeline*, a system that represents changes in global temperatures over the 20th century: horizontal axis represents time, and color represents temperature, with bluer shades representing cooler periods and redder shades representing warmer periods:

FIGURE 6: A timeline of global temperatures.

Suppose the timeline is purely analog (it lacks any textual labels). Position on the horizontal axis is locative; two points on the line cannot share the same horizontal position. In effect, climate timelines are pictures of temperature over time. Syntactically, it's analogous to a one-dimensional photograph. But its interpretation function maps space to time and color to temperature. This content is represented iconically despite being non-spatial, non-visual, and relatively abstract. Similar points apply to audio recordings, which use time, instead of space, to represent time. These

[21] For spatial domain-specificity, see Greenberg [2021] and Kulvicki [2020]. For visual domain-specificity, see Hopkins [1998] and Gregory [2010]. For non-abstract domain specificity, see Quilty-Dunn [2020] and Rivadulla-Duró [forthcoming]. Relatedly, while the relation between the literature on *depiction* and iconic representation is complex, many paradigmatic depictive representations, including paintings and photographs, are clearly iconic. In the depiction literature, it is common to maintain that depiction is a constitutively visual phenomenon (Hopkins [1998, 2005], Lopes [1996], Neander [1987], Schier [1986], Wollheim [1998]).

[22] Analogous restrictions on the vehicles (rather than contents) of iconic representation are similarly unmotivated. For example, audio recordings make use of vehicles that are neither visual nor spatial. Nevertheless, by locating sounds in time (and representing both analogically), on our view they belong to the same representational kind as photographs that locate colors in space.

examples illustrate why we should be cautious about drawing substantive conclusions about the content of a representation merely on the basis of its iconic format.

Locative structure serves an important representational need by establishing constraints on the contents a system can express. Consider what would happen if spatial position in *Color Photographs* were qualitative, rather than locative. If two pixels could share spatial position, and thus represent the same worldly spatial location, then the very same location could be represented as being both red and green. But there is no need to represent such a content, since no actual object can be both red and green in the same spot. By making spatial position locative, we preclude the possibility of photographs with these sorts of incoherent contents, thereby allowing the system to better serve our needs and purposes. For many representational systems, there will be similar constraints on how features that are represented can be combined. Whenever one constructs a system whose syntactic rules mirror these semantic constraints, one will end up with a system that has locative structure.

§3 Iconic Representation as Analog Locative Structure

Our theory of iconic representation has three components:

MOLECULARITY:	Iconic representations have multiple parts.
ANALOGICITY:	Iconic representations represent analogically.
LOCATIVITY:	Iconic representations have locative structure.

In what follows, we argue for each of these components and show how they result in a satisfying and informative theory of iconic representation.

3.1 Molecularity

We started with a discussion of the parts principle, one of the most influential claims in philosophical work on iconic representation. The main motivation behind the parts principle is that iconic representations can be decomposed into successively smaller parts that represent successively

smaller parts of the content. This motivation presupposes that iconic representations have a rich mereological structure.

The condition that iconic representations are molecular can be motivated by comparing atomic and molecular representations that meet our other two conditions on iconicity. Contrast a mercury thermometer with a temperature map that uses hues from blue to red to represent temperature. Both the thermometer and the temperature map represent analogically; both have a syntactic feature (height or color) that mirrors the semantic feature it represents (temperature in both cases). Moreover, both the thermometer and the map have locative structure: every part of the map occupies a unique spatial position, and every part of the mercury thermometer (namely, the improper part corresponding to the whole thermometer) has a unique value with respect to height.[23] But there remains an important sense in which the map is a "picture" of temperature, whereas the mercury thermometer is not. The difference is that the map can be decomposed into many syntactic parts while the thermometer consists of a single atomic syntactic part. Our theory delivers the verdict that the map is iconic, while the thermometer is not.

3.2 Analogicity

Consider a representational system that is molecular and locative but fails to be analog. Let *World Cup Qualification* be a representational system in which national flags each lie adjacent to three numbers. The flags represent the national football team of the corresponding nation. The numbers represent the win-loss-draw record of that team in the qualifying rounds.

[23] This is the limit case of locative structure. See *fn.*19.

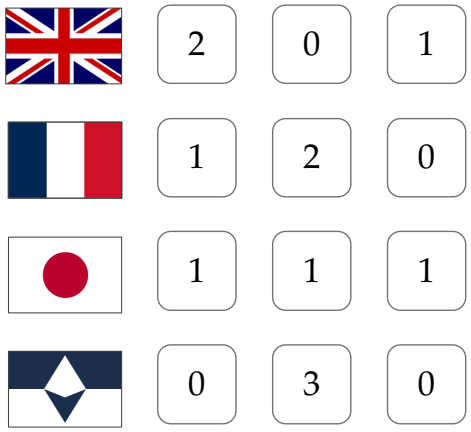

FIGURE 7: A World Cup qualification score board.

This system has multiple parts (corresponding to the different flags and numbers). Moreover, flags are locative because no two teams can share their flag. The trio of numbers is qualitative because two teams can share the record of wins, losses, and draws. But even though *World Cup Qualification* is molecular and locative, it's not iconic. The problem is obvious: all of its features represent symbolically, rather than analogically. Numbers and flags are paradigms of symbolic representation.[24]

As another example, recall *Seating Chart*, which uses a spatial array of names to represent where each student sits in a classroom. Unlike *World Cup Qualification*, *Seating Chart* is partially analog. While the names are symbolic, *Seating Chart* uses space to represent analogically. However, *Seating Chart* also fails to be iconic. The reason is that it's merely partially, rather than fully, analog. A representation must have exclusively analog features in order to count as iconic.

3.3 Locativity

Consider again *Box of Thermometers*, where a box containing a collection of thermometers represents the daily temperatures over a summer. The representation has multiple parts: the individual thermometers. Those

[24] Flags are symbolic despite being two-dimensional colored arrays and thus bearing some resemblance to paradigmatic iconic representations like photographs and paintings. This is because there is no structural correspondence between flags and the countries they represent. See Lee, Myers, & Rabin [2023] and Giardino & Greenberg [2015: 2].

thermometers have only one feature: the height of a vertical bar, which represents analogically. However, height is qualitative, because two thermometers can share the same height value. On our view, *Box of Thermometers* is not an iconic representational system precisely because it lacks locative structure. By contrast, *Color Photographs*, the representational system to which all color photographs belong and which often serves as the paradigm of iconic representation, has both qualitative (color) and locative (spatial position) features.

There may be a temptation to think that *Box of Thermometers* is not iconic merely because it's too simple, rather than because it lacks a locative structure. But consider a more sophisticated version of this representational system, which we call *Fancy Box of Thermometers.* Suppose each thermometer is colored a shade of blue, with the degree of saturation representing humidity and the degree of brightness representing degree of cloud coverage.

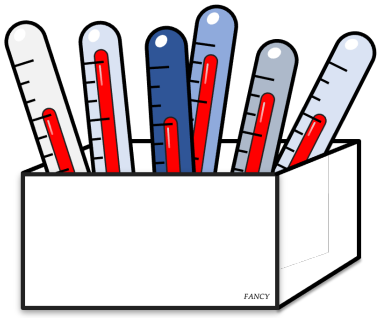

FIGURE 8: A fancy box of thermometers.

We might even suppose that the number of possible configurations of the fancy thermometers is identical to the number of possible configurations of the color, *x*-axis, and *y*-axis features of a color photograph. Given this, *Fancy Box of Thermometers* and *Color Photographs* will be systems of roughly equal complexity.

Nonetheless, there remains an important difference between the systems: *Color Photographs* has locative structure, since each pixel has a unique spatial position, whereas *Fancy Box of Thermometers* has no locative structure, since it utilizes only qualitative features. Two thermometers can share the height value of their vertical bar, width value of horizontal bar,

and color value (i.e. they represent the same temperature, humidity, and cloud coverage), whereas two individual pixels of a color photograph cannot share both their *x*- and *y*- spatial position values. Despite its additional complexity, *Fancy Box of Thermometers* remains a system of mere aggregates; it doesn't organize parts into a locative structure.[25]

3.4 Analog Locative Structures

We are now in position to state our theory of iconic representation: iconic representations are **analog locative structures**. They are molecular in that they have multiple parts. They are locative in that the parts have unique values with respect to some features. And they are analog in that they involve structural correspondence between vehicle and content. These three simple criteria provide a satisfying and informative theory of iconic representation.

The components of our theory accommodate some central ideas that drive much prior work on iconic representation. The first idea is that iconic representations are molecular rather than atomic. Photographs, paintings, and dioramas are complex representations made up of many parts that each contribute to the meaning of the whole. This is a common starting point for many theories of iconic representation, including our own. The second idea is that iconic representations are structurally similar to what they represent.[26] Photographs and sculptures are structurally similar to the scenes or objects that they depict. Our theory explains and accommodates this idea with the analogicity requirement. The third idea is that iconic representations are, in some sense, spatial.[27] Photographs and sculptures

[25] Quilty-Dunn [2020] and Quilty-Dunn & Green [2021] endorse a necessary condition on iconic representation that they call 'holism': every part represents multiple properties at once. For example, every part of a color photograph represents both space and color. These authors take the parts principle and the holism principle together to distinguish the iconic from the discursive. However, *Fancy Box of Thermometers* satisfies both principles: every part represents temperature, humidity, and cloud coverage at once, and every part represents a part of the total set of temperatures, humidities, and cloud coverages that is represented. Nevertheless, *Fancy Box of Thermometers* fails to be iconic.

[26] See Burge [2018].

[27] See Kosslyn [1981] and Clarke [2022].

are spatially organized and depict spatially organized contents. While our theory denies that iconic representation must make use of physical space, it accommodates the intuition by appealing to the more general notion of locative structure. Locative structure can be manifested by position in physical space but can also be manifested by other kinds of features. This allows our theory to accommodate representations such as audio recordings that are intuitively iconic but are temporally rather than spatially structured.

Our discussion began with the parts principle, which holds that iconic representations are those where parts of vehicles represent parts of contents. The parts principle is on the right track. It captures the fact that iconic representations are molecular. But it's too weak. To be iconic, the parts need to both represent in the right way (analogically rather than symbolically) and be structured in the right way (locatively rather than qualitatively). Many representations have multiple parts, yet don't satisfy these two further criteria. Some, like *Seating Chart*, aren't iconic because the parts represent symbolically. Others, like *Box of Thermometers*, aren't iconic because the parts are mere aggregates. Only when a representation meets all three conditions does it fall within the realm of the iconic.

3.5 Degrees of Iconicity

Our theory focuses on defining conditions that are individually necessary and jointly sufficient for iconicity. But it's possible to use our framework to accommodate the idea that iconicity comes in degrees. Since our theory involves three ingredients—MOLECULARITY, ANALOGICITY, and LOCATIVITY—one might aim to analyze degrees of iconicity in terms of those ingredients. We think there are various ways of precisifying degrees of iconicity, and for limits of space we won't explore these options in detail.[28]

[28] In our view, degrees of analogicity provide the most natural ingredient for developing a degree-theoretic notion of iconicity. One option is to appeal to the proportion of features that represent analogically. Another option is to appeal to individual features, determined by the extent to which each feature involves structural correspondence between vehicle and content. See Lee, Myers, & Rabin [2023] for an in-depth account of degrees of analogicity. See Greenberg [2023] for an alternative account.

If iconicity comes in degrees, then we face the question of how much iconicity a system needs before counting as iconic *simpliciter*. One option is **maximal:** to be iconic simpliciter is to have a maximal degree of iconicity. The other option is **non-maximal:** to be iconic simpliciter is to surpass some intermediate degree of iconicity. For example, maps that involve both analog and symbolic features will satisfy the non-maximal definition but not the maximal definition.

Our own inclination is to use 'iconic' in the maximal way. More specifically, we prefer a usage of 'iconic' on which iconicity requires a representation to have *only* analog features. This definition better accords with our own patterns of usage of 'iconic'. In addition, it enables the term 'iconic' to capture the difference between maps (with symbolic elements) and photographs. However, all the substantive components of our theory would still stand if we were to choose instead to use 'iconic' in a non-maximal way.

3.6 A Methodological Point

Our principal goal in this paper is to identify an interesting representational kind, rather than to analyze ordinary usage of the word 'iconic'. In other words, this paper is more an endeavor in conceptual engineering than conceptual analysis. At various points, we appealed to the reader's sense for whether a given representation ought to count as iconic. These moves are intended to illustrate which representational systems are similar or different in kind, rather than to track how people use the word 'iconic'. We then aim for a theory of iconic representation that maps joints between representational kinds in a way that captures these similarities and differences.

For some readers, there may be divergence between intuitions about which representations count as iconic and the predictions of our theory. But we take that to be compatible with the main explanatory task of this paper, so long as one also agrees that we have identified some interesting and fruitful distinctions between representational kinds.

Over the rest of the paper, we showcase various ways in which our theory is explanatorily fruitful. In §4, we apply our theory to iconic

representations with more unruly representational structures. In §5, we apply our theory to mental representations and argue that our theory clarifies a long-standing debate in cognitive science about the notion of a "functional space." In §6, we explain how the notion of locative structure is useful for thinking about even non-iconic representational systems.

§4 Beyond Paradigm Systems

When we defined 'locative structure', we invoked several simplifying assumptions:

ATOMICITY:	Representations are wholly composed from atomic parts.
FEATURE COMPLETENESS:	Every part has a value along every feature.
COMPOSITIONALITY:	The features of molecular representations are wholly determined by the features of and relations between their parts.

These assumptions enabled us to focus on a class of "well-behaved" representational systems. This in turn allowed us to develop a streamlined version of our core theory. We now explain how to generalize our theory—and especially the notion of locative structure—when we relinquish these assumptions. This is an important virtue of our account, since most theories of iconic representation focus on paradigms of iconic representations, which nearly always satisfy the assumptions above. Because of this, they apply only to a restricted class of representational systems, and aren't naturally generalizable to these more unruly systems.

4.1 Atomless Systems

A system that fails ATOMICITY will be at least partially *gunky*, in that it has some syntactic parts that always have further syntactic parts.[29] One might

[29] The gunkiness, or infinite decomposability, comes at the level of the syntactic rather than physical structure of the representation. There is thus no assumption here that the physical

think, for example, that paintings or film photographs have no atomic parts. Instead, perhaps these representations are composed of spatial regions, colored in various ways, that always have subregions.

Our prior discussion of locative structure focused only on systems with atomic parts. The atomic parts of *Color Photographs* were individual pixels. If we can no longer assume ATOMICITY, then we need to revisit how the notion of a locative feature works. One function the restriction to atomic parts played was to partition a complex representation into a collection of parts that together compose the whole representation. However, it turns out that the important aspect was the partition, not the atoms. So long as we can slice the system into parts—atomic or molecular—that compose the whole, we can use the already introduced apparatus of §2.1 to determine which features are locative and qualitative, relative to that partition. We then declare a feature locative *simpliciter* (in that representational system), when it comes out as locative on every possible partition of the system.

Let's walk through this procedure in more detail. Let a **partition** of a representation be a division of the whole representation into distinct parts that (*a*) together compose the whole and (*b*) do not overlap. The elements of the partition can be atoms, molecules, or a mix of the two. One way to partition a color photograph is to slice it up into 2×2 blocks of pixels. In the general case, elements of the partition can be molecular parts; these parts will take multiple values along a given feature. For example, each 2×2 block will take multiple spatial positions and multiple colors as its value on the spatial position and color features (namely, the spatial positions and the colors of the block's individual pixels).

Importantly, even if the 2×2 blocks have no atomic parts, because the parts each have further parts *ad infinitum*, the block can still take as its spatial position value exactly the spatial positions it occupies. This is a straightforward and intuitive way of handling feature values for molecular parts. If we instead think of the system as operating with separate x-coordinate and y-coordinate features, then each block will be associated

item (e.g., an oil painting) is infinitely decomposable into smaller and smaller physical parts.

with a set of *x*-coordinate *y*-coordinate pairs: in each case exactly the pairs whose spatial positions the block occupies.

Intuitively, it matters how much of each value a molecule has. A 2×2 block with three green pixels and one red pixel has a different color value than a block with three red pixels and one green. For systems that satisfy ATOMICITY, this can be measured simply by counting the number of atoms in the molecule that take a certain value. For systems that don't satisfy ATOMICITY, one would need to appeal to a different way of measuring the amount of each feature value had by each molecule.[30]

If a color photograph is partitioned into 2×2 blocks, spatial position will be locative relative to that partition if and only if each element of the partition (each block) has a unique spatial position value. In *Color Photographs*, each block does because each 2×2 block has a unique set of spatial positions. Again, this is true regardless of whether we think of spatial position as a single feature or break it into separate *x*-coordinate and *y*-coordinate features. In the latter scenario, each 2×2 block will be associated with a unique set of *x*-coordinate *y*-coordinate pairs. As in §2.1, *x*-coordinate and *y*-coordinate will form a locative set, and both *x*-coordinate and *y*-coordinate will count as locative features.[31] This result will repeat itself on every possible partition. Therefore, in the *Color Photographs* system, spatial position remains locative even when the assumption of atomic parts is relaxed. On the partition into 2×2 blocks, two distinct 2×2 blocks can have the same color values. This will occur when all the sub-parts of the 2×2 blocks have the same color values. This entails that color is qualitative relative to this partition, and thus qualitative in the representational system.

This procedure—partitioning the complex representations of the system in all possible ways, checking the locativity of features on each partition, and then declaring the feature locative when it passes on every

[30] A natural choice for continuous systems is the Lebesgue measure. See Tao [2011].

[31] As noted in §2.1, the locative set must be minimal. The fact that each 2×2 block has a unique set of *x*-coordinate, *y*-coordinate, and hue triples does not make hue locative, because *x*-coordinate and *y*-coordinate achieve this result alone.

possible partition—is how we handle systems that lack a privileged partition into atomic parts.[32]

Some have thought that iconic, unlike discursive, representations lack *canonical decomposition*.[33] We think it's plausible that the framework developed here will be useful for handling worries about representations that lack canonical decomposition. Each partition captures a different way that a system can be decomposed into parts. So, the suggestion we make for handling systems with no privileged atomic parts—checking all the partitions—should also work for handling systems with no canonical decomposition.

4.2 Feature Incomplete Systems

If a representational system doesn't satisfy FEATURE COMPLETENESS, then some parts of the system will lack some representationally relevant features. For example, consider a board game where most pieces occupy positions on a map, but where some special pieces lie outside of the map and hence lack a value with respect to map position. We characterized locative features as the features of a system on which every part takes a unique combination of values. For feature incomplete systems such as the board game, one might worry that the fact that two pieces lack map position values will render map position no longer locative.

The natural solution is to **conditionalize** the definition of 'locative'. In particular, define a set of features as *locative* just in case each part *that instantiates all those features* has a unique combination of values along those features. If some parts fail to take a value along a feature, then there's no violation yet of that feature counting as locative. By consequence, even if a board game contains multiple pieces that lack spatial position, it could still turn out that spatial position is locative in that board game.

---

[32] It's possible that a representational system with atomic parts might have other additional partitions that are privileged. For such systems, one could consider modifying the procedure discussed here by evaluating whether features come out as locative on all privileged partitions. For simplicity, we ignore such situations here.

[33] See Kosslyn [1980], Fodor [2007], [2008], and Quilty-Dunn [2016], [2020].

This approach captures the spirit of locative structure and the notion of a feature value. Values are specific ways that a feature can be instantiated. To fail to have a color is not to have color in a specific way. We have also spoken about locative features in terms of parts not sharing feature values. The approach of treating two parts that lack a feature altogether as failing to share a feature value is common sense: if neither you nor I have a favorite curling team, we don't share a favorite curling team. All of this entails that one should not automatically interpret any part that fails to instantiate a representationally relevant feature as taking a "null value" on that feature.[34]

4.3 Non-Compositional Systems

If a representational system doesn't satisfy COMPOSITIONALITY, then it will have some *emergent* features, meaning features of molecules that aren't determined by features (including relational features) of their proper parts.[35]

Consider, for example, a museum wall display of digital photographs of animals, each with the same image resolution, but where images of more massive animals are larger. For example, the image of a whale and the image of a wombat both have a 3840 × 2160 resolution, but the whale image is much bigger than the wombat image. Importantly, the size of an image isn't determined by the feature values (colors and spatial positions) of the parts of that image. In fact, the color and spatial positions of the pixels can be held fixed even while varying the size of the image. Therefore, image size counts as an emergent feature of this system.

[34] We leave open the possibility that in some cases, a representational system could operate using "null values". Additionally, though the language is similar, the approach here should not be construed as taking sides in the debate between Kulvicki [2015] and Rescorla [2009] on "the absence intuition", in which the notion of a null value plays a role. See also Camp [2007].

[35] The discussion here may remind some readers of holistic phenomena where features of a part are determined by how it relates to other parts. But these systems don't generate a problem for our view. Whether the feature values of atomic parts are determined holistically, atomistically, or otherwise, the atomic parts still have feature values, and we can apply our apparatus to determine whether a given feature is locative or qualitative.

Our theory has a straightforward way of applying the definition of 'locative' to emergent features. Such cases can be interpreted as a special kind of violation of feature incompleteness: in non-compositional systems, there are parts—including all atomic parts—that lack values along the emergent features. The solution is to apply the definitions of 'locative' and 'qualitative' to molecular parts (alongside atomic parts). More specifically, we can deploy the conditionalization procedure mentioned previously, which we used when dealing with failures of feature completeness.

Let's apply this to the animal images example. To determine whether image size is locative vs. qualitative, we check whether the parts that have a value on the size feature (meaning whole images) can share their value. For this system, it's natural to think of size as qualitative, since two animals can have the same mass. However, if image size represented something different—such as a strict ordinal ranking of one's favorite animals—then image size might instead come out as locative. This approach will work for determining the locative vs. qualitative status of any emergent feature.

4.4 Beyond Paradigm Systems

We appealed to two moves in explaining how to apply the notion of 'locative feature' when ATOMICITY, FEATURE COMPLETENESS, or COMPOSITIONALITY fail. The first was partitioning: we appealed to partitions over whole representations. This enabled us to handle systems that are either gunky or that lack canonical decompositions. The second move was conditionalization: we restricted the relevant parts to those that have values along the features under question. This enabled us to handle feature-incomplete and non-compositional systems.

It's worth noting that neither of these moves requires substantive modification of our original definition of 'locative feature'. Here again is that definition: a feature is **locative** for a system just in case it's a member of a **locative set**, meaning a minimal set of features such that within every representation of the system, each part takes a unique combination of values on those features. The definition is already compatible with partitioning because it invokes parts in general, not just atomic parts. To make the definition compatible with conditionalization, we need only

restrict the quantifier 'each' to parts that have values along the relevant features.

In our view, most discussions of iconic representation have focused on narrow classes of representational systems, with specific kinds of representational structures. A general aim of this paper has been to focus on a wider range of representational systems in order to better identify which properties are essential to iconicity and which ones merely incidental to its paradigms. A theory of iconic representation ought to be general enough to handle the rich variety across representational kinds, yet specific enough to identify the properties that unify the class we call 'iconic'.

Most discussions of iconic representation don't investigate systems that fail to satisfy ATOMICITY, FEATURE COMPLETENESS, and/or COMPOSITIONALITY. On top of that, it's not easy to develop a theory that can handle these types of systems. A significant virtue of our theory is that it can accommodate these representational phenomena.

§5 Iconic Mental Representation

Much of the interest in iconic representation has stemmed from a desire to understand mental representations, such as perceptual states and mental images. At the same time, many theorists have been skeptical that mental representations could be meaningfully similar to pictures and sculptures. In this section, we explain how our account of iconic representation can clarify the widely-invoked notion of functional space and be fruitfully applied to mental representations.

5.1 The Functional Space Problem

Many have thought that physical space plays a particularly important role in iconic representation. Kosslyn [1981] writes that iconic representations must occur in a "spatial medium," and Clarke [2022] writes that iconic representations "have a fundamentally spatial structure."[36] But there's an obvious problem for theories of iconic representation that privilege

[36] An obvious motivation for thinking that iconic representations must be spatial comes from resemblance theories of depiction which hold that iconic representations must resemble the spatial scenes that they depict. See Peirce [1982] and Hyman [2012].

physical space: they make it nearly impossible for mental representations to be iconic. It's exceedingly unlikely that there are spatially extended pictures carved into our brain tissue.[37] More generally, the mere fact that spatial distances between neurons are representationally irrelevant shouldn't preclude mental representations from counting as iconic.

Many authors, including those quoted above, respond to this problem by replacing the notion of physical space with that of **functional space**.[38] This is the standard move by philosophers and cognitive scientists who favor the parts principle but who recognize that individuating parts via physical space will fail to render mental representations iconic. The basic idea is that iconic mental representations rely on functional analogues of physical space. This move allegedly enables iconic mental representations to function as if they are spatially extended without literally being spatially extended.

However, discussions of functional space typically provide only cursory explanations of what it means for a representational system to be functionally spatial.[39] In most cases, the concept of functional space is simply taken for granted. For this reason, some authors have decried the notion as spurious: Pylyshyn [2002: 167] says that the notion of a functional

[37] Some authors, such as Kosslyn *et al*. [2006: 100], point to the retinotopic organization of the primary visual cortex to argue that these areas of the brain "use space on the cortex to represent space in the world." But even if some iconic mental representations happen to be spatially organized, this isn't essential to their status as iconic.

[38] Kosslyn [1981, 1994] is the canonical defender of functional space. More recently, Quilty-Dunn [2020: 813-4] argues that mental images have "functional analogues of the spatial properties of the represented scene," and therefore have a "functional–spatial structure." Similarly, Clarke [2022: 568] gives an account of mental imagery according to which "depicted spatial relations are mirrored by the spatial relatedness of nodes in a functional space, and the nodes within this functional space mirror the feature values they depict."

[39] Kosslyn [1994: 5] is an exception. He writes that "a depictive representation is a type of picture, which specifies the locations and values of configurations of points in a space…The space in which the points appear need not be physical, such as on this page, but can be like an array in a computer, which specifies spatial relations purely functionally. That is, the physical locations in the computer of each point in an array are not themselves arranged in an array; it is only by virtue of how this information is "read" and processed that it comes to function as if it were arranged into an array (with some points being close, some far, some falling along a diagonal, and so on)." See also Tye [1990: §3.1].

space "is merely metaphoric and not explanatory," while Beck [2018: 4] says that "it is far from clear just what a functional space is supposed to be, or how it might be rigorously characterized."[40] Let the **functional space problem** be the challenge of explicating what it is for a representational system to have a functional space.

The problem may be framed as a dilemma. On the first horn, the notion of a functional space is just that of a physical space. This option makes the notion of a functional space too restrictive. For example, it excludes representations such as mental images and perceptual experiences (which don't rely on physical space) from the realm of the iconic. But these are exactly the kinds of mental representations that motivate the invocation of functional space in the first place.

On the second horn, a functional space is anything that can be modeled by a mathematical space (meaning a set paired with some relations). But this second option makes the notion of a functional space too permissive. For example, the parts of *Box of Thermometers* (the individual thermometers) can be modeled in a one-dimensional state space where location in the space is determined by mercury height and the elements of the space bear "taller than" relations to each other.[41] But no one would want to count *Box of Thermometers* as functionally spatial.

The ideas developed in this paper give us the tools to resolve this dilemma. Here's our analysis of when a representational system has a functional space:

> a representational system has a **functional space** :=
> the system has a set of features that are both locative and analog.

The locative requirement ensures that each part occupies a unique position in the feature space. This ensures that functional spaces retain one of the

[40] Fodor [2003: 36-7] expresses additional skepticism in the same vein.

[41] Making the thermometers representationally richer and generating a two (or more) dimensional state space won't help. *Fancy Box of Thermometers* is exactly such a system. But it's no more iconic than *Box of Thermometers* is. In the language invoked earlier, both are "mere aggregates."

key roles associated with physical space. Just as points in physical space are unique, so too are points in functional space. The analog requirement precludes systems that involve no structural correspondence between vehicles and contents (such as sentences and ordered lists). This ensures that only systems with the right kind of structure count as having functional spaces.[42]

More specifically, functional spaces instantiate counterparts of spatial relations. Consider a system that analogically represents metric structure. The vehicles of this system will have parts standing in metric relations that are functionally analogous to spatial distance, with some parts "closer together" and others "further apart." However, these relations may be, for example, relations of functional connectedness between neurons, rather than literal spatial distances.[43]

The reader might wonder what the difference is between our theory of iconic representation and our theory of functional space. After all, both appeal to analog representation and locative structure. Here's the difference: iconicity requires that *all* features represent analogically, whereas a system's having a functional space requires only that the *locative* features represent analogically. Because of this, our definition of 'functional space' allows for the possibility of non-iconic representations that have functional spaces. In other words, our analyses entail that having a

[42] One could choose to add further requirements for having a functional space. First, one might require that functional spaces have metric or topological structure. Second, one might require that the system use the features that generate the functional space to represent actual physical space as their content. Third, one might choose to rule out systems without multiple parts from counting as functionally spatial. Most cases discussed in the literature on functional space meet these additional requirements. For example, all of Kosslyn's examples involve representation of physical space. But it isn't clear that Kosslyn thinks that in order to count as having or using functional space, the system *must* represent physical space. It's far from clear that the best, most unified, and/or explanatorily elegant theory of functional space incorporates any of these further addenda. For both simplicity and theoretical scope, we choose to not build in any of these suggestions into our account.

[43] The term "functional space" stems from the idea that a system can count as spatial in virtue of the functional roles that it plays in the mind. Our account respects this motivation insofar as a set of features can count as locative and analog in virtue of the functional roles those features play.

functional space is a necessary, but not sufficient, condition for iconic representation.

To motivate this, consider again *Seating Chart*. This representation is non-iconic (since the names represent symbolically), but still functionally spatial (since its locative feature—spatial position—represents analogically). As another example, consider *Color Word Photograph*, in which each pixel of the photograph contains a color word ('vermillion', 'chartreuse') rather than a color. This system is likewise non-iconic (since each word represents symbolically), yet also functionally spatial (since its locative feature–spatial position–represents analogically).[44]

Our approach offers the tools to solve the functional space problem by identifying the representational role that physical space plays in photographs, paintings, and other paradigm iconic representations. Our theory of iconic representation is built around that role (locative structure), rather than around a common realizer of that role (physical space). On our view, physical space is one step removed from the true marker of iconic representation: it's merely one of many ways of manifesting an analog locative structure.

At the same time, it's unsurprising that many paradigms of iconic representation make use of physical space. This is because using physical space as a representational feature is a particularly natural way to generate an analog locative structure. On the syntactic side, it's often natural to think of distinct parts as corresponding to distinct spatial positions. On the semantic side, there's rarely a need to represent distinct objects as occupying the same spatial position. Given this, our theory can account for why many authors have felt that physical space is important for iconic representation while still explaining how mental representations such as mental images and perceptual experiences can count as iconic.[45]

---

[44] Many maps have this kind of structure: they have both a functional space that represents analogically as well as elements that represent symbolically (such as symbols for different kinds of buildings).

[45] Our suggested solution to the functional space problem should be useful even for theorists who don't explicitly invoke functional space or the parts principle. For example, Clarke [2022] claims that iconic representations are "analog maps." But he says little about what a map is, other than that a map is spatial. Since his explanatory target is pre-attentive

5.2 How Mental States Represent Iconically

Our solution to the functional space problem delivers the welcome verdict that a representation can be iconic even if it doesn't use physical space as a representational feature. This means that there needn't be miniature pictures carved into the brain in order for a mental representation to be iconic. However, one might still reasonably wonder how exactly to apply our account of iconic representation to mental representation. Some might think we need fundamentally different theories to explain mental representations, such as mental images and perceptual experiences, vs. non-mental representations, such as photographs, sculptures, and topographical maps.[46]

The ingredients that comprise our theory of iconicity—molecularity, analogicity, and locativity—can be had by both mental and non-mental representations. First, both public and mental representations can have multiple parts (where the parts are individuated by the syntax of the system). The parts of a mental representation might be symbols in the language of thought or positions in the visual phenomenal field; the parts of a non-mental representation might be pixels or audio clips. Second, both mental and non-mental representations can be analog. For mental representations, the structural correspondence might be between neural firing rates and luminance magnitudes; for non-mental representations, the structural correspondence might be between spatial relations on a canvas and spatial relations in the world. Third, both mental and non-mental representations can have locative structure. In both cases, the non-shareability of feature values might be either due to the physical properties of the vehicles themselves, or the functional architecture of the representational system.[47]

To further illustrate, consider a toy example. Consider a network of neurons whose parts are neuron-clusters, where each cluster contains three

---

vision, it's unlikely that he intends 'spatial' to include only physical space. A natural way to precisify the relevant sense of 'spatial' is to appeal to locative analog structure.

[46] See Coelho Mollo & Vernazzanni [2023] for a statement of this worry.

[47] See Lande [2021] for a related idea in the context of mental representations.

neurons. The firing rate of the first neuron represents *x*-coordinate position, the firing rate of the second neuron represents *y*-coordinate position, and the firing rate of the third neuron represents luminance. We can further suppose that each of these features represents analogically, with firing rate structurally corresponding to either spatial or luminance dimensions. In effect, this is a neural analog of a black and white photograph. As in a photograph, *x*- and *y*-coordinate position will be locative just in case two parts cannot share the same values along both dimensions.

Of course, it's physically possible for two neurons to fire at the same rate. However, it may be syntactically impossible for a mental state to represent two parts sharing the *x*- and *y*-coordinate positions. If the consumer system is unable to process patterns of activity in which two sets of *x*- and *y*- neurons share the same firing rate, then this is reason to think that the representation is not syntactically well-formed and therefore that neural firing rate is a locative feature. As an analogy, many strings of words are physically possible but syntactically ill-formed.

Things get more complicated when we move to real-life examples. *Place cells* in the hippocampus are neurons that fire when the organism occupies specific locations in space.[48] This has led many cognitive scientists to refer to them as 'cognitive maps.' However, place cells are not spatially arranged in ways that mirror their spatial content. Instead, place cells stand in coactivation relations to each other; activation in one place cell will tend to cause activity in place cells that represent nearby locations.[49] Because co-activation relations structurally correspond to spatial relations, it's plausible that co-activation relations analogically represent space. Moreover, if two place cells cannot stand in the same coactivation relations to other place cells, then coactivation relations will be locative. If these conditions are met, then cognitive maps will instantiate a functional space (and be iconic). This captures the idea that coactivation relations in cognitive maps play the representational role that spatial distance plays in physical maps.

[48] See O'Keefe & Dostrovsky [1971] and Grieves & Jeffery [2017].

[49] See Diba & Buzsáki [2007] and Dragoi & Tonegawa [2013].

§6 Representational Joints

It's widely recognized that the analog vs. symbolic distinction and the atomic vs. molecular distinction denote two representational joints. By contrast, the joint between locative vs. non-locative has been underappreciated. In fact, the concept of locative structure can illustrate similarities between analog and symbolic representations, and highlight important differences across molecular representations.

Consider the following sequences of representations. The top row features, from left to right, a single pixel, a collection of paint swatches, and a picture. The bottom row features a word, a grocery list, and a sentence:

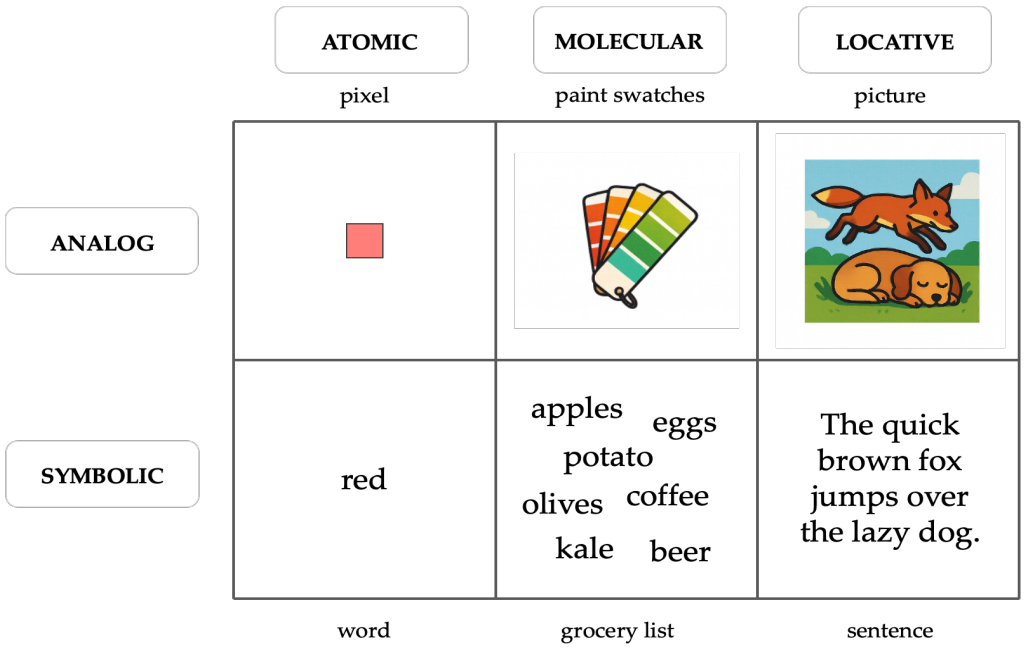

FIGURE 9: A table illustrating some representational kinds.

Since the top row is analog while the bottom row is symbolic, it's obvious that there are important representational differences across the rows. But there's also an underlying similarity within each column. Pixels stand to paint swatches stand to photographs as words stand to unordered lists stand to sentences. As we move across each row from left to right, we move from atoms to mere aggregates to locative structures.

Many discussions of representational format collapse distinctions between types of molecular representations. However, just as there are important reasons to distinguish between unordered lists and sentences in the case of language, so too is there reason to distinguish between mere aggregates and locatively structured representations within the realm of the analog. The ring of paint swatches you find at the hardware store represents an unordered set of colors. But photographs do more. They don't merely represent an aggregate of colors; they also locate those colors in a structure.

While our focus has been on the role of locative structure as an ingredient for iconic representation, it's worth noting that locative structure is likewise useful for thinking about other kinds of representational systems, including natural language. In natural language, position in a syntactic tree is a locative feature: two lexical items cannot occupy the same position in a syntactic tree. By contrast, position in an unordered list is representationally irrelevant. While lists are mere aggregates, sentences are more than the sum of their parts.

The notion of locative structure is also useful for thinking about *hybrid systems*, or systems with some analog features and some symbolic features. For example, consider the difference between maps and diagrams. Most maps have spatial features that represent analogically and locatively and symbols that represent symbolically and qualitatively (the same symbol can appear at multiple spatial positions). By contrast, most diagrams function in the converse way. In a pie chart, the labels are symbolic and locative, while the sizes of the "slices" are analog and qualitative (since two slices can be the same size).[50] This illustrates how the notion of locative structure enables us to draw natural distinctions within the class of hybrid systems, corresponding to whether the analog or symbolic features are locative.

The fact that the concept of locative structure is useful for thinking about a variety of representational systems—even non-iconic ones—is

[50] These claims about maps and diagrams aren't exceptionless. For example, *Pinned Map* is cartographic but has a locative feature (namely, pin color) that represents symbolically. However, these exceptions are compatible with our claim that our distinctions provide fruitful ways of carving up representational kinds and comparing the representational structures of particular maps and particular diagrams.

evidence that it denotes a natural representational kind. Furthermore, the generality of the notion—defined simply in terms of whether or not features are sharable—means that it's applicable, in principle, to any representational system.

Our theory also clarifies the contested relationship between the iconic and the analog. On our approach, analog representation is a necessary condition for iconic representation. To be iconic, a representational system must represent analogically, but it also must be both molecular and locative. The iconic, therefore, is a subset of the analog.

Conclusion

Iconic representations—pictures, drawings, sculptures, dioramas, audio recordings, mental images, and so forth—represent in a distinctive way. One respect in which they represent distinctively is that they represent analogically, rather than symbolically. Another respect is that they are complex: they have multiple parts. These parts form structured collections, rather than mere aggregates. This structure is provided by the locative features of the representational system. When there are no locative features, there is no such structure. The idea that iconic representations are molecular and analog is relatively familiar. But the notion of locative structure is novel to the iconic representation literature.

We started with an examination of the parts principle. We argued that no version of the parts principle yielded a wholly satisfactory analysis of iconic representation. But the iterations of the parts principle did collectively point towards a promising idea: iconic representations are, in some sense, structured collections of analog representations. This served as a guiding idea for our theory, according to which iconic representations are locatively structured analog representations.

Alongside developing a positive theory of iconic representation, we have developed conceptual tools that are useful for theorizing about representational systems and kinds more generally. The notion of locative structure enables us to define, in more systematic terms, what it is for a system to have a functional space. The categories of the locative, qualitative, analog, and symbolic yield a fruitful taxonomy of representational kinds.

They allow us to identify salient similarities as well as differences across quite distinct representational phenomena, including photographs, pie charts, and sentences.

A noteworthy feature of our theory is that it focuses on the representational role that spatial position plays in representations like pictures and photographs, rather than on the realizer of that role. By doing this, our theory is able to identify the commonalities between iconic representations that make use of physical space (such as pictures and photographs) and iconic representations that use other features to generate locative structure (such as audio recordings and mental images).

We began this paper with a familiar adage: "A picture is worth a thousand words." When thinking about iconic representation, it may be tempting to extend the thought: "A picture is worth a thousand pixels." But a picture is more than the sum of its pixels. To be iconic, a representation must also organize its parts into the right kind of structure. That structure, we have argued, is locative structure. A picture isn't merely a collection of pixels: it's an analog locative structure.

References

Abell, Catharine (2009). Canny resemblance. *Philosophical Review* 118 (2):183-223.

Balog, Katalin (2009). Jerry Fodor on Non-conceptual Content. *Synthese* 167 (3):311 - 320.

Beck, J. (2018). Analog mental representation. *WIREs Cognitive Science, 9*(6), 1–10.

Beck, Jacob (2019). Perception is Analog: The Argument from Weber's Law. *Journal of Philosophy* 116 (6):319-349.

Burge, Tyler (2018). Iconic Representation: Maps, Pictures, and Perception. In Wuppuluri Shyam & Francisco Antonio Dorio, *The Map and the Territory: Exploring the Foundations of Science, Thought and Reality*. Springer. pp. 79-100.

Camp, Elisabeth (2007). Thinking with Maps. *Philosophical Perspectives* 21(1), 145–182.

Carey, Susan (2009). *The origin of concepts*. New York: Oxford University Press.

Casati, Roberto & Varzi, Achille C. (1999). *Parts and Places: The Structures of Spatial Representation*. MIT Press.

Clark, Austen (ed.) (2000). *A Theory of Sentience*. New York: Oxford University Press.

Clarke, Sam (2020). Beyond the icon: Core cognition and the bounds of perception. *Mind and Language* 37 (1):94-113.

Clarke, Sam (2022). Mapping the Visual Icon. *Philosophical Quarterly* 72 (3):552-577.

Coelho Mollo, Dimitri & Vernazzani, Alfredo (2023). The Formats of Cognitive Representation: A Computational Account. *Philosophy of Science* (3):682-701.

Diba, Kamran, and György Buzsáki.2007. "Forward and ReverseHippocampal Place-Cell Sequences during Ripples." Nature Neuroscience 10 (10):1241–42.

Dragoi, George, and Susumu Tonegawa. 2013. "Distinct Preplay of Multiple Novel Spatial Experiences in the Rat." Proceedings of the National Academy of Sciences 110 (22):9100–5.

Fodor, Jerry (2007). The revenge of the given. *Contemporary Debates in Philosophy of Mind*:105–116.

Fodor, Jerry A. (2003). *Hume Variations*. Oxford, GB: Oxford University Press.

Fodor, Jerry A. (2008). *Lot 2: The Language of Thought Revisited*. New York: Oxford University Press. Edited by Jerry A. Fodor.

Green, E. J. & Quilty-Dunn, Jake (2021). What Is an Object File? *British Journal for the*

Philosophy of Science 72 (3):665-699.

Greenberg, Gabriel (2021). Semantics of Pictorial Space. *Review of Philosophy and Psychology* 1 (4):847-887.

Greenberg, Gabriel (2023). The Iconic-Symbolic Spectrum. *Philosophical Review* 132 (4):579-627.

Gregory, Dominic (2010). Pictures, pictorial contents and vision. *British Journal of Aesthetics* 50 (1):15-32.

Grieves, Roddy M. and Kate J. Jeffery. (2017). "The Representation of Space in the Brain." Behavioural Processes 135:113–31.

Hopkins, Robert (1998). *Picture, Image and Experience: A Philosophical Inquiry*. New York: Cambridge University Press.

Hopkins, Robert (2005). "The speaking image: visual communication and the nature of depiction." In Mathew Kieran, *Contemporary Debates in Aesthetics and the Philosophy of Art*. Malden, MA: Wiley-Blackwell. pp. 135--159.

Hyman, John (2012). "Depiction", *Royal Institute of Philosophy Supplement*, 71: 129–150.

King, Jeffrey C. (2007). *The nature and structure of content*. New York: Oxford University Press.

Kosslyn, Stephen M. (1980). *Image and Mind*. Harvard University Press.

Kosslyn, Stephen M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review* 88 (1):46-66.

Kosslyn, Stephen M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. MIT Press.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery.* Oxford University Press.

Kuhn and Aristodemo. (2017). Pluractionality, iconicity, and scope in French Sign Language. *Semantics and Pragmatics*, 10(6), 1-49.

Kulvicki, John (2015). Analog Representation and the Parts Principle. *Review of Philosophy and Psychology* 6 (1):165-180.

Kulvicki, John (2015). Maps, Pictures, and Predication. *Ergo: An Open Access Journal of Philosophy* 2(7): 149-174.

Kulvicki, John (2020). *Modeling the Meanings of Pictures: Depiction and the Philosophy of Language*. Oxford, GB: Oxford University Press.

Lande, Kevin (2021). Mental Structures. *Noûs* (3):649-677.

Langland-Hassan, Peter (2025). The Imagery Debate Exhumed and Reanimated. *The British Journal for the Philosophy of Science*.

Lopes, Dominic (1996). *Understanding Pictures*. Oxford: Clarendon Press.

Lee, Andrew Y. ; Myers, Joshua & Rabin, Gabriel Oak (2023). The Structure of Analog Representation. *Noûs* 57 (1):209-237.

Maley, Corey J. (2011). Analog and digital, continuous and discrete. *Philosophical Studies* 155 (1):117-131.

Neander, Karen (1987). "Pictorial representation: a matter of resemblance." *British Journal of Aesthetics* 27(3): 213–26.

O'Keefe, John and Jonathan Dostrovsky (1971). "The Hippocampus as a Spatial Map." Brain Research 34: 171–75.

Parsons, Josh (2007). Theories of Location. In Dean Zimmerman, *Oxford Studies in Metaphysics: Volume 3*. Oxford University Press UK.

Peacocke, Christopher (2019). *The Primacy of Metaphysics*. New York, NY, United States of America: Oxford University Press.

Peirce, Charles S. (1982). *The Writings of Charles S. Peirce: A Chronological Edition*, eds. M. Fisch, C. Kloesel, E. Moore, N. Houser et al., Bloomington IN: Indiana University Press.

Pylyshyn, Zenon W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences* 25 (2):157-182.

Quilty-Dunn, Jake (2016). Iconicity and the Format of Perception. *Journal of Consciousness Studies* 23 (3-4):255-263.

Quilty-Dunn, J. (2017). Syntax and Semantics of Perceptual Representation. Ph.D. dissertation, The Graduate Center, City University of New York.

Quilty-Dunn, Jake (2020). Is Iconic Memory Iconic? *Philosophy and Phenomenological Research* 101 (3):660-682.

Quilty-Dunn, Jake (2020). Perceptual Pluralism. *Noûs* 54 (4):807-838.

Rescorla, Michael (2009). Predication and Cartographic Representation. *Synthese* 169(1): 175–200.

Rivadulla-Duró, Andrea (forthcoming). Iconic Prioritization and Representational

Silence in Emotion. *Australasian Journal of Philosophy*.

Salmon, Nathan U. (1986). *Frege's Puzzle (2nd edition)*. Atascadero, CA: Ridgeview Publishing Company.

Schier, Flint (1986). *Deeper Into Pictures: An Essay on Pictorial Representation*. New York: Cambridge University Press.

Schlenker, Philippe ; Lamberton, Jonathan & Santoro, Mirko (2013). Iconic variables. *Linguistics and Philosophy* 36 (2):91-149.

Soames, Scott (1987). Direct Reference, Propositional Attitudes, and Semantic Content. *Philosophical Topics* 15 (1):47-87.

Sober, Elliott (1976). Mental representations. *Synthese* 33:101-48.

Tao, Terence (2011). An Introduction to Measure Theory. Providence, RI: American Mathematical Society.

Tye, Michael (1991). *The Imagery Debate*. Cambridge: MIT Press.

Wollheim, Richard (1998). On pictorial representation. *Journal of Aesthetics and Art Criticism* 56(3): 217–26.